Measuring Quality in Initial Teacher Education

A Literature Review for Scotland's MQuITE Study

By Emilee Rauschenberger, Paul Adams, & Aileen Kennedy

Table of Contents

1. Introduction: Measuring Quality in ITE

1.1. Purpose of the Literature Review: MQuITE Study in Scotland

The MQuITE project seeks to develop and implement a context-appropriate approach to measuring quality in initial teacher education. As a result, the focus of this literature review is on the various conceptualisations and measures of "quality" vis-à-vis teaching, teachers, and initial teacher education (ITE) across the globe. As Scotland considers what types of programmes and routes best prepare teachers for success in the classroom, there is a glaring need for a collaboratively-created framework with which to measure the quality of teacher preparation programmes. Such a framework would not only provide data to inform external evaluation and internal improvement but also provide a coherent and consistent model with which to compare the strengths, weaknesses, and innovative features of the various pathways into teaching.

This is a conceptual, not systematic, review of the literature on ITE and quality. A conceptual review differs from a systematic review in its purpose and focus. While a systematic review typically entails researchers uncovering all evidence relevant to "a very specific empirical question, often posed in a cause-and-effect form" (Kennedy 2007, p.139), a conceptual review aims to gain new insights into an issue. As a result, conceptual reviews do not focus on what is known, empirically, about a defined problem. Instead, they seek to (1) uncover what is not known or unexplored in the literature, (2) understand how others in the past have studied or viewed a problem, (3) discover the intertwining of a problem with other issues, or (4) consider the nature of questions or methodologies with which an area has been researched. Thus, conceptual reviews come in a variety of forms and may be integrative, theoretical, methodological, or historical (Kennedy, 2007). The review here examines the issue of quality ITE; thus it considers how others have defined and attempted to measure quality in ITE, issues related to this challenge, and unexplored yet important elements of ITE that affect quality.

The purpose of this literature review is to offer three types of insights. First, the discussions here offer insight into the dimensions of international debates regarding the quality of initial teacher education, and how various stakeholders propose measuring quality in ITE. The political contexts at the local, national, and international level combine with cultural and historical peculiarities to shape education generally, and ITE

in particular, in different parts of the world. Hence, it is important to understand how similar terms, ideas, or philosophies can be interpreted and enacted in idiosyncratic ways in context. Secondly, the review offers insight into a number of tools and frameworks currently being used to measure teacher and teacher education quality in various localities. The strengths and weakness of these strategies and models are reviewed as well as the methodological and practical challenges faced by researchers in implementing them. Thirdly, the review of the literature also identifies measurements that have been developed as well as others that are potentially needed. These are areas within ITE that are seldom addressed in existing quality measures or frameworks. Yet, there are programmes and studies that suggest that such aspects affect the quality of ITE and the effectiveness of the teacher graduates they produce. Thus, the literature review incorporates discussions of the current debates, contextually-derived tools and frameworks, and the gaps of measurement that are critically important to consider when attempting to devise one's own model of quality indicators.

1.2 Defining Key Terms

To facilitate a fruitful discussion of the literature on teacher quality and effectiveness of ITE, it is useful to designate what is meant by key terms that are commonly used in the field of initial teacher preparation and debates regarding education quality. First, the MQuITE study focuses on *initial* teacher preparation programmes in Scotland to date, and thus this review often employs the commonly used term "initial teacher education", or simply ITE, to refer to this university-based pathway into teaching which includes both four-year bachelor degrees programmes and one-year post-graduate programmes and a new two-year Masters programme. By the latter half of the 20th century, the majority of initial teacher preparation programmes in countries worldwide were located in higher education institutions. However, as mass education rapidly expanded, the need for more new teachers quickly outpaced supply in most contexts. The need for additional teachers spurred the introduction of alternative pathways in some countries, starting with the U.S. and England in the 1980s and 1990s respectively. In some countries, notably England, ITE has been conceptualised as initial teacher training (ITT) to signify its instrumental purpose in providing practical skills and knowledge for effective teaching and reflect the introduction of school-led and

employment-based teacher preparation programmes (Furlong *et al.*, 2000). In the English context, that has meant a further shift in the language used to describe teacher preparation programmes and student teachers, which are now commonly referred to as ITT "providers" and "trainees" respectively. However, for this review, ITE is consistently used to refer to university-based training courses as well as alternative pathways into teaching, most of which utilize higher education institutions to some degree to educate emerging teachers.

"Quality", and relatedly, "quality measurement" or "indicators" are other key terms that have many synonyms. The term "quality" is defined by the Oxford Dictionary as "the standard of something as measured against other things of a similar kind; the degree of excellence of something". However, the meaning of the term varies widely across local and national contexts. In the literature, "quality" is most often used inter-changeably with "effectiveness". Thus, a quality (or rather "high-quality") teacher is one that is most "effective". "Effective" in doing *what* exactly depends on the cultural and political context but typically refers to the ability of teachers, or teacher educators, to enable students to understand and master course content and taught skills. Still, quality or the effectiveness of a teacher or an ITE programme can also refer to the cultivation of wider moral values and orientations within students. Meanwhile, quality measurement, or indicators, are terms used to describe how conceptual understandings of quality are measures in different contexts. Frequently used synonyms for this idea, which commonly appear in the literature, include "assessment", "accountability" measures, and "evaluations". Yet, again, the meaning of these various terms and synonyms subtly changes with the type of study conducted or context in which it is used.

1.3 Scope and Methodology of the Review

The following literature review focuses on ways of conceptualizing and measuring the quality, namely the impact and effects, of ITE programmes. To maintain this core focus as well as be relevant to prevailing contexts of the time, the review was limited in terms of its subject area, time period, and language. The review covered studies that focused on either (1) ITE programmes, their components and effectiveness, in preparing teachers as measured in various contexts, and/or (2) teacher effectiveness as related to ITE and related routes into teaching. Since the meaning of terms relating to quality in

the education sector shift in different contexts, close attention was paid to the cultural and political environment and historical trajectories in which studies of teacher education programmes were situated and how quality indicators developed in relation to their local contexts.

In terms of time period, only sources published during or after 2005 were considered. The decision to limit the review to literature published in the past 12 years was made for the purpose of timeliness and relevance. Doing so enabled the review to focus on current practices and policies in the field related to the most recent political, economic, and social developments in national and international contexts. Educational contexts have significantly shifted in the past two decades, particularly in the U.K. and the U.S. The devolution of powers to home countries within the U.K. in 1997 and the passage of the federal legislation *No Child Left Behind* in the U.S. in 2001 dramatically altered educational landscapes in terms of school policy and management. Given the time lapse between when research is conducted and when results are typically reported, it was agreed that studies from 2005 onward best captured this changing environment and its impact on initial teacher education. Furthermore, the global economic crash of 2007-8 and the ensuing recession also significantly shaped policies pertaining to state schools and higher education over the past decade. Since then, government funding for public services has been drastically cut across European countries in the name of austerity in hopes of aiding ailing economies. Given these shifts in the contexts affecting ITE, the review of the literature focuses on studies published in the past twelve years.

Although the rationale regarding the time scale has referenced developments in mainly Western countries, the geographical scope of the literature is global. Studies in a variety of countries and settings were considered. However, the review only examined studies and documents written in English.

The types of studies examined in this review varied and included (1) academic studies published in peer-reviewed journals, (2) books or book chapters, (3) reports from government agencies, think tanks, and non-profit groups with consideration of the political affiliations and ties to the publishing organization, and (4) published or unpublished doctoral theses (including PhDs and EdDs). Additional sources examined included related articles from reputable media outlets and other forms of policy literature (i.e., government papers, hearing transcripts, conference papers, etc.). As recommended by Kennedy (2007), literature reviews that include a wider scope of materials than

simply peer-reviewed academic journals benefit from a wider ideological and epistemological perspective.

Using these parameters, the literature for the review was obtained by searching the GoogleScholar, Education Resources Information Center (ERIC), and DiscoverEd databases (the academic search engine for the University of Edinburgh). Search terms included "quality" and frequently used synonyms for quality – such as assessment, evaluation, effectiveness, and so forth – in conjunction with initial teacher education, preparation or training. In addition to database searches, the bibliographies of numerous other literature reviews and policy analyses in this area were reviewed and entire journals whose domains encompassed this general topic were also searched. Despite these procedures and rules, many complications arose during the literature search, most having to do with what constitutes quality measures and how to treat studies of teacher quality measures that were often intertwined with considerations of the quality of ITE.

1.4 Structure of the Literature Review

The structure of the review is as follows. In the next section, current global trends in education policy and teacher preparation are discussed in order to highlight the importance of the topic at hand and help explain why evidence and accountability has become a main theme in education generally. A look at the unique policy context and ITE sector in Scotland complements the global perspective and captures the current context in which the MQuITE study will be conducted. After considering both the global and Scottish contexts, a review of the literature on ITE locates this review in the research landscape. In the third section, the review focuses on the general approaches with which others in the literature have conceptualized and assessed ITE quality, or effectiveness. The components of ITE that are typically assumed to contribute to its quality are examined and the ways in which they have been measured are discussed along with competing orientations toward evaluation studies. In the fourth section, different approaches and studies of the quality of ITE are examined in details and differences among them are highlighted. In the fifth and six sections, the findings of the review are synthesised and reflected upon, particularly in connection with issues of teacher recruitment and retention. In addition, gaps in the literature identified and discussed before concluding thoughts are summarised.

2. Current Contexts and Trends in the Field of ITE

2.1 Raising Teacher Quality (and by extension ITE Quality) as a Global Priority

Across the literature on education, measuring teacher effectiveness and the role of ITE have become a global focus. Politicians, school leaders, and university faculty alike are grappling with ways to measure and prove (or refute) the impact of ITE on the quality of the teachers they produce (Wineburg, 2006). This focus has been the result of the latest paradigm shift in national debates and the global educational discourse, which have come to a general consensus that quality teachers are the most important factor in pupil achievement.

The focus on teachers, teaching quality, and ITE represents the latest thinking about how to raise educational achievement on a societal scale. As Cochran-Smith and Villegas (2015) highlight, schools, teacher education and research on both, have been shaped by larger social, political, and economic forces and ideologies, particularly in the U.S., over the past fifty years. They highlight that, in developed countries, the shift away from a manufacturing-based economy to a more global knowledge-based one gave rise to new labour markets, new patterns of production and consumption, and worldwide mass migration. Accompanying these developments was the rise of neo-liberalism, an ideological perspective whereby individualism, free markets, and private goods are prioritized over other goals. These developments, along with neoliberal ideas, shaped education policy and purposes, leading to three distinct trends witnessed in the late twentieth century: (1) changing conceptions of how people learn and what they need to know in a knowledge economy, (2) increasingly diverse student population and growing school inequality, and (3) unprecedented attention to teacher quality and accountability (Cochran-Smith and Villegas, 2015, p. 9).

These trends – representing socio-political, ideological, demographic, and/or intellectual developments – have affected the views of teachers and the role of, and research on, teacher education across developed countries. In contrast to the post-World War II period in which teaching was seen as a craft carried about by rather autonomous professionals, these global developments have led to teachers and education generally being positioned as more important politically, economically and socially. As a result, governments, the media, and the public have become more involved in educational

debates, fuelling more and more studies of teachers, their effectiveness, and their preparation.

In 2005, the Organization for Economic Cooperation and Development (OECD) published the findings from its study of teacher policies in 25 countries. The report, entitled *Teachers Matter* (OECD, 2005), analysed the expansive research on determinants of pupil learning and concluded that the most important influence on pupil learning – which is "potentially open to policy influence" – was teachers and teaching, or simply put, "teacher quality" (p. 2). The report acknowledged that the largest greatest variation in student outcomes is attributable to students' social background, motivation, and abilities, but focused on the how countries could attract, develop, and retain high-quality teachers more effectively. This policy thinking was further spread by reports by McKinsey consultants whose across-countries studies of school systems suggested the quality of teachers, rather than more funding or smaller class size, was a key lever for raising pupil achievement (Barber and Mourshed, 2007; Barber, Mourshed and Whelan, 2007; Mourshed, Chijioke and Barber, 2010).

Although countries have arrived at the "teacher quality" paradigm from different pathways that reflect their own social and political histories, governments across the world have come to view education as a national resource and engine of countries' economic growth and competitiveness (OECD, 2005). It is worth nothing, however, that there is some evidence contradicting the assumption that educational achievement is a good indicator of and/or contributor to national economic growth and development. As Ramirez et al. (2006) found, although student achievement may be positively correlated with a country's direct spending on education and the overall strength of the economy, this relationship is weak and not backed up by other measures of economic productivity and development. Similarly, data from UNESCO and TIMSS indicate that public expenditure on education as a percentage of GNP, GDP growth, and gross domestic investment *are not* closely associated with student achievement while public expenditure on education per capita and GNP per capita *are* (Wiseman, 2007). This analysis suggests that education should not be the sole, or primary, means through which governments plan to grow their economy.

Despite such evidence, the logic of linking a more knowledgeable future workforce with greater economic prosperity and competitiveness persists. As the U.K. government's White Paper, *The Importance of Teaching* (2010), states:

What really matters is how we are doing compared with our international competitors. That is what will define our economic growth and our country's future. The truth is, at the moment we are standing still while others race past.

(DfE 2010, p. 3)

Thus, education – specifically teacher quality as the main determinant of student achievement – is considered the prime means through which to compete and thrive as a nation economically.

Meanwhile, as demands for teacher quality have increased, so has the need for greater quantity of teachers. With pupil enrolments on the rise and a generation of teachers nearing retirement age, the need for more teachers has grown. In developed countries, education policy initiatives such as class size reduction have only compounded that need while many less developed countries are expanding schooling and struggling to certify new teachers and those already in the field.

2.2 The Changing Role, Identity, and Responsibility of the Teacher

In addition to the global consensus that teachers matter most, the expectations and responsibilities of teachers have also changed. As the OECD report (2005) and other studies highlight, teachers are now expected to assume much broader roles. In the classroom, teachers are expected to create a multi-cultural learning environment, an inter-disciplinary curriculum and take into account the ability and development of individual students, including pupils with additional support needs. At the school levels, teachers are expected to work collaboratively, create systems of evaluation and data tracking to inform planning, and use technology in teaching effectively, among other things. On a wider community and professional level, teachers are expected to advise parents, build links within the community, and improve their own professional skills and knowledge through ongoing professional development and in carrying out teacher-led research (Jyrhämä *et al.*, 2008; Valcke, 2013).

Furthermore, additional expectations and pressure have been exerted on teachers working within low-income communities to close the "achievement gap" – a term referring to the persistent difference in school achievement and college attendance rates between the poor, mainly minority pupils and their more wealthy and predominantly

white counterparts (Harris and Herrington, 2006; Ladson-Billings, 2006; Reardon, 2013). In Scotland, this is conceptualised as the 'attainment gap', and is related principally to children living in poverty. The belief that teachers are "miracle workers" as suggested by the U.S. President Obama's Education Secretary (Goldstein, 2009) who can, through education, "trump poverty" (Kopp, 2011, p. 75) is an underlying current within the "teacher quality" policy discussions. This is despite persistent data indicating that teachers have only a limited influence on student achievement compared to the impact of non-school factors (OECD, 2005; Chetty, Friedman and Rockoff, 2014).

While expectations and responsibilities for teachers have multiplied, demographic shifts in student populations nationally and globally have made teaching arguably more challenging. English-language learners, minorities, and students with additional support needs now make up a larger portion of students in many classrooms, especially in urban areas. This development presents new challenges for teachers, who must grapple with the question of how best to support such a diverse population of students. In this context, the need to understand and measure the quality and effectiveness of ITE in preparing effective future teachers has become critically important.

2.3 From "Standards and Accountability" to GERM in Education and ITE

Since the 1980s, the drive to raise the quality of education nationally, mainly in terms of tests scores, resulted in five main policy and reform principles. Sahlberg (2015) identifies these as *competition* (principle 1) among schools which, accompanied by the idea of giving parents more *choice* (principle 2) regarding which school their child attended, would drive up performance. Consequently, *standardization* (principle 3) in the 1990s was spurred on by the popularity of outcomes-based education in the 1980s. This resulted in standards-driven education policies during the 1990s, initially in Anglo-Saxon countries. Standardization refocused attention on the outcomes of schooling, namely student learning and schools' performance. While curriculum standards directed the context of schooling, teaching standards sought to define what "good teaching" was and what it looked like, though such standards varied from locality to locality. Common to all types, however, was the inherent belief, or assumption, among policy-makers and education reformers that "setting clear and sufficiently high

performance standards for schools, teachers, and students will necessarily improve the quality of desired outcomes" (Sahlberg, 2011, p. 177).

In addition to competition, choice, and standardization, the standards and accountability movement in the 1990s and 2000s focused attention on *core subjects* (principle 4) in the schools curriculum including literacy, numeracy, and the natural sciences. International testing (i.e., PISA and TIMSS) only increased attention on these subjects. Finally, these developments led to widespread calls for and use of *test-based accountability* (principle 5) to hold educators to set standards, evaluate pupil progress in core subjects, and provide data to parents and schools to encourage choice and competition respectively. For the U.K., this particular approach to education reform started with Margaret Thatcher's philosophy based on competition, information, and choice. However, these ideas were embraced internationally, first by Canada, the U.S. and Australia and then further afield through think tanks, the World Bank, and other international organisations. It is worth acknowledging here, though, that Scottish education policy and legislation has remained separate from that of the UK since 1885 (see more detailed discussion in the following section).

Meanwhile, these principles fed the paradigm of "evidence-based" policy-making, which called for "scientific evidence" on which to base education policy decision-making. Wiseman (2010) highlights that this call for empirical evidence of effective practice to guide policy-making rests on the common belief among policy-makers and education reformers that "school knowledge is abstract and universal" and that "empirical evidence is an efficient indicator of knowledge and learning" (p.1). As a result, calls for evidence have mainly reflected more positivistic perspectives and epistemologies, as traditionally favoured in the natural sciences rather than research orientations of those in the social sciences. Such calls for "scientific evidence" on which to base policy have shaped the type of educational research conducted and used to inform policy-makers.

For example, in the U.S., the What Works Clearinghouse (WWC), established in 2002 with funding from the U.S. Department of Education, develops summaries of the scientific evidence for "what works" in education. Yet, the WWC went against the advice of social scientists and rejected traditional views and methods of summarising and drawing conclusions from research. Instead, the WWC frequently makes judgments of "best practice" through discounting the cumulative body of social research in favour

of fewer studies with specific criteria (Stockard and Wood, 2016). In addition, the WWC criteria favour quantitative and experimental research designs and randomized field trials for producing scientific evidence on the effectiveness of educational programmes (Chatterji, 2005). Meanwhile, in England, government calls for research that can directly inform policy-making have pressured researchers to adopt applied or practice-based approaches in their studies of educational practice and programmes (Whitty, 2006).

The call for evidence of best practice, together with the principles of the "standards and accountability movement", have spread, evolved, and become so ingrained in educational discourse around the world that it has been renamed by some as the Global Education and Reform Movement, or GERM (Conway, 2013; Sahlberg, 2016). In addition to continuing to intensify policy focus on the five tenets of the standards and accountability movement, GERM also "assumes that the most effective way to improve educational systems is to bring well-developed innovations to schools and classrooms from outside, often from the business world" (Sahlberg, 2016, p. 137). For school teachers and ITE faculty, this has translated into the call for the *systematic* training of teachers and staff, or the recruitment of "better and smarter" people into schools. Despite national education success stories (i.e. Finland) that run counter to the market-based thinking of GERM, these new accountability expectations and the emerging policy focus on teacher quality and ITE spread to many other parts of the globe by of the end of 1990s (Cochran-Smith and Villegas, 2015).

The standards and accountability movement in education has also had an impact on the ITE sector. In the U.S., Cochran-Smith (2005) highlights how, prior to the mid 1990s, the focus in ITE was on processes – (1) how student teachers learned to teach, (2) how social and organizational contexts affected such learning, (3) what knowledge base supported that learning and the development of effective practice, and (4) how teacher candidates' beliefs and attitudes evolved (or did not) over time. However, with the rise of the standards and accountability movement in education, Cochran-Smith (2005) argues that a "new teacher education" is emerging with emphasis primarily on outcomes, not process. This new conceptualization is one in which ITE is "constructed as a public policy problem, based on research and evidence, and driven by outcomes" (*ibid*., p. 3) – a development which Cochran-Smith argues has positive and negative implications but which also presents dangers and opportunities. These developments in

the U.S. have parallels in other parts of the world, most notably countries such as England (Furlong *et al.*, 2000) and Australia (Mayer, Pecheone and Merino, 2013).

These developments and debates around ITE are common in developed countries and have led, in some nations, to a curtailing of the role of universities in the education of teachers (Roth, 1999). This has stemmed chronic shortages of teachers and a questioning of the effectiveness of ITE, both of which have prompted the creation of additional routes into teaching and a relaxation of entry standards for individuals entering the profession. In the U.S, England, Australia, and parts of Europe, there is an on-going struggle between those who see the need to strengthen ITE and the knowledge base of teachers and those who favour more employment-based, learning-on-the-job approaches.

In summary, the shift from a focus on ITE inputs and process, to principles of accountability, standards, and evidence have led to an increase in research and questions regarding the current role and impact of ITE. Thus, the need for studies and measurement systems of the effectiveness and quality of ITE programmes continues to intensify.

2.4 The Scottish Position: Policy and ITE Contexts

While global trends are important to take into account when considering policy and research designs, local contexts are equally (if not more) important. The context of Scotland is the vantage point from which international research and wider trends are examined to inform the MQuITE study. As such, Scotland's unique political and ITE contexts are important to review here.

Scottish education policy has been developed separately from the UK more widely since 1885 and the creation of the Scottish Office, but since the creation of the Scottish Executive (now called the Scottish Government) in 1999, the story of Scotland and Scottish education has been more resolutely one of devolution and independence. However, there are contrary positions for devolution and its relationship with the welfare state: on the one hand are positions that suggest that devolution offers the opportunity to enhance social democracy; on the other, that it might lower levels of welfare provision (Scott and Wright, 2012). It is against this backdrop that readings of social policy positions should be made.

What is notable since devolution in 1999 is that successive administrations have sought to distance themselves politically and ideologically from administrations in England (Watson, 2010). Since devolution, policy differences seem to have become more entrenched between Scotland and U.K., reflecting a more collective ideology on the part of the former (Scott and Wright, 2012) with inclusion as a central good of the system and education at the heart of this endeavour. Over time many have argued that devolved government is the way to get to the heart of the social justice issue (Scott and Wright, 2012) and that Scottish issues are such that social justice needs to be at the heart of Scottish social policy (Mooney and Scott, 2012). Additionally, the drive to address poverty and social exclusion has become markedly "Scottish" in orientation in the years following devolution with increased willingness to blame Westminster priorities for the failure to tackle issues (Scott and Wright, 2012). Foregrounded here are myths about Scotland and the Scottish people. Part of the "Scottish myth" is that there exists a democratic, egalitarian system (Raffe, 2004), central to which "…are claims that Scotland (and 'the Scots') is imbued with a series of progressive values, notable among which is a sense of, and commitment to, egalitarianism, collectivism and social justice" (Mooney and Scott, 2012, p. 2). It is plausible to suggest that in general, public support for public services is higher in Scotland than in England and there are less-discernible desires for private or for-profit systems of delivery. This is borne out in figures about the relative proportion of expenditure on the public sector in Scotland compared with that in England (Cairney and McGarvey, 2013).

Central to the Scottish National Party (SNP) minority government position from 2007 were three positions for policy-making (Sanderson, 2011). First, a shift to co-production of policy with local government was promoted. A key feature of the "new" politics in Scotland, and a direct result of the ways in which elections for the parliament were organised in terms of a desire for no overall control by one party, has been co-operation and partnership. In 2008, the Scottish Government described a concordat with local government to deliver a more collaborative agenda. Second, a recognition of the potential for greater experimentation within Scotland under devolution was highlighted. In this context, SNP policies marked a clear departure from New Labour's target setting regime. Finally, the importance of evidence and analysis was emphasised. The SNP shifted attention from micro-managing delivery to improving performance, and hence

there was a need to strengthen the evidence base upon which performance could be measured (Sanderson, 2011).

In order to understand better the effectiveness and impact of policy in relation to outcomes, evaluation was noted for its potential to inform future policy development and resource decisions (Sanderson, 2011). A shared agenda was set that contrasted with previous central-surveillance techniques and the project was to be socially democratic with a distinctive Scottish accent with fairness married to wealth as defining issues of the day (Arnott and Ozga, 2010a).

But economic prosperity and social justice sit in uneasy tension as gains in one area might be construed as lapses in the other. Despite this, there were two clear, mutually reinforcing positions in policy discourses (Arnott and Ozga, 2010b). First, the economy: competitiveness was referenced "outwards" by comparison with other states in terms of globalised economies. Here the SNP sought to position Scotland with small, democratic states such as those in Scandinavia or the Baltic region: globalisation was a scripting feature of Scottish education policy. Referencing was also "inwards" by allusion to the ways in which Scotland has the ways and means to ensure it can remain competitive through harnessing its natural talent. For Arnott and Ozga (2010a) this amounted to the use of "modernised nationalism" whereby "nation", as already implicit in much education discourse, was mobilised through a "simultaneous process of 'inward' referencing of ideas of fairness and equality, combined with 'outward referencing' which places Scotland in alignment with new comparators in education, mainly in the Nordic and Baltic states" (p. 347). The second position concerned flourishing: an emphasis on community, fairness and inclusivity. Here the education agenda was one of achieving greater autonomy, and all policies acted within a context of moves towards independence (Arnott and Ozga, 2012).

Alongside this more collaborative approach to governing at national and local level, the discourse of partnership has become increasingly prominent in Scottish education policy development, with a growing emphasis on (1) stakeholder involvement, (2) a more prominent role for pressure groups, and (3) significant public dialogue between stakeholders, government ministers and civil servants (Hulme and Kennedy, 2016). This partnership approach to policy development became even more explicit in teacher education with the publication of a major review of teacher education (Donaldson, 2011), which mandated greater partnership working between schools, local

authorities and universities as a "panacea" (Kennedy and Doherty, 2012) for teacher education, particularly in the early phase (ITE and induction).

The wholesale review of teacher education, commissioned by the Scottish Government in 2010, was led by recently retired head of Her Majesty's Inspectorate of Education, Graham Donaldson. The report, entitled *Teaching Scotland's Future,* was published in early 2011 and recommended 50 changes to teacher education from pre-entry to headship, including the recommendation that:

> In line with emerging developments across Scotland's universities, the traditional BEd degree should be phased out and replaced with degrees which combine in-depth academic study in areas beyond education with professional studies and development. These new degrees should involve staff and departments beyond those in schools of education.
>
> (Donaldson, 2011, p. 40)

This recommendation, of all the Donaldson recommendations, arguably required the most significant re-deployment of resources. Duly, six of the eight Scottish providers of ITE designed new undergraduate (UG) degrees. Of the other two providers, the University of the Highlands and Islands currently only offers postgraduate ITE, while the University of Stirling already offered undergraduate ITE compliant with the Donaldson recommendation.

The evidence base for this move, recommendation 11, is somewhat opaque however. Indeed, the report asserts that "Overall, the evidence gathered by the Review, indicates that recently-qualified teachers, probationers and students are generally satisfied with their experience in initial teacher education and induction" (Donaldson, 2011, p. 34). Despite no formal evidence to support the changes, Donaldson (2011, p. 40) asserts that "Degrees which concurrently combine significant academic study outwith education with rigorous professional development for teaching offer a more relevant way forward than the traditional BEd programmes'. In the 2016 intake, UG primary students accounted for 710 of the 3230 ITE places in Scottish universities.

Significantly, although these undergraduate primary programmes have undergone significant reform, Scottish ITE as a whole now has a more explicit emphasis on partnership between schools, local authorities and universities as a result of recommendation 15:

> New and strengthened models of partnership among universities, local authorities, schools and individual teachers need to be developed. These partnerships should be based on jointly agreed principles and involve shared responsibility for key areas of teacher education.'
>
> (Donaldson, 2011, p. 91)

Subsequent to these Donaldson-related reforms, the Scottish Government developed the *National Improvement Framework for Scottish Education* (Scottish Government, 2016b) and, under the new Cabinet Secretary for Education and Skills, a delivery plan for education (Scottish Government, 2016a). These documents provide a mandate for ITE providers to think more creatively about how their provision can meet a number of pertinent challenges. As such, the Cabinet Secretary, John Swinney, called on university providers to consider how, as a sector, they could address the following priorities:

1. Increased numbers of teachers in shortage subjects such as STEM and Home Economics

2. Teachers who can work between primary and secondary sectors to support the transition phase

3. Increased opportunities for specialism within the primary workforce, e.g. STEM and modern languages

4. PGDE and induction year combined more coherently, and potentially over a shorter timescale

5. Opportunities for teachers to complete a full Masters degree during ITE or the induction year

6. Increased availability of distance or work-based routes into teaching

7. Increased numbers of black and minority ethnic teachers, and of male teachers in the primary sector

8. Increased numbers of teachers able to teach through the medium of Gaelic

On 1st December 2016 the Cabinet Secretary announced formally his Government's support for an innovative package of new routes, further diversifying the portfolio of ITE routes in Scotland, with a twin focus on producing both more, and better, teachers. This package includes a two-year Masters offering transition qualification (nursery – secondary 3 generalist pathway or primary 5 – secondary 6 subject specialist pathway), a specialist Maths teaching route spanning primary and early

secondary phases, partnership delivery with STEM-focused university departments, and, controversially, a 'fast-track' route which combines the PGDE and induction year.

The announcement of these new routes was quickly followed by the establishment of a teacher recruitment campaign, "Teaching Makes People", specifically targeting graduates in shortage subjects (particularly STEM subjects) (see http://teachinscotland.scot). This new campaign occurred – despite the radical plans already in place – in response to mounting concerns about specific teacher workforce shortages. During the first half of 2017, increasing concerns about teacher workforce availability in specific subjects and in specific geographical locations in Scotland became so intense that the Scottish Parliament Education and Skills Committee launched an enquiry, resulting in ITE coming under increasing scrutiny and critique.

During this period of significant change for ITE in Scotland, the former Scottish Teacher Education Committee (STEC), comprising the Heads of School or Deans of each of the Scottish universities offering teacher education, reformed itself as the Scottish Council of Deans of Education (SCDE) (www.scde.ac.uk). Sensing a potential threat to the long-held presumption that ITE in Scotland should be university-based, one of the first acts of the newly (re)established body was to commission Professor Ian Menter to write a literature review on "the role and contribution of higher education in contemporary teacher education" (Menter, 2017). Coincidentally, in the week that Menter's report was formally launched, the Cabinet Secretary for Education, John Swinney, announced his intention to seek bids to develop a "Teach First" type route in response to severe recruitment challenges and teacher shortages. This announcement has been met with some resistance, not least from the dominant teacher union in Scotland, the Educational Institute of Scotland (EIS), who on 10th June passed an emergency motion "condemning the reported intention of the Scottish Government to tender for new approaches to Initial Teacher Education (ITE) which would potentially bypass universities" (EIS, 2017). The EIS stated that it is "not opposed to alternative pathways into teaching – we are opposed to shortcuts which would impact on the high standards of our teaching profession" (EIS, 2017).

So, at the time of writing, ITE in Scotland is in a state of significant flux, which brings with it exciting possibilities to break from the shackles of historically established narrow and conservative routes, but at the same time, the threat of dilution of ITE in response to undeniably pressing workforce demands. The need to agree and implement

a sector-wide framework of measuring the quality of this increasingly diverse range of ITE routes has never been more pressing.

2.5 A Glance at the Research on Initial Teacher Education as a Field of Study

Turning away from global and national context and toward the terrain of the literature, it is worth noting that the study of ITE is a relatively new research field compared to other more established branches of social science. As a result, theoretical and methodological break-throughs have been gradual and slow to surface (Borko, Liston and Whitcomb, 2007). In their survey of the empirical literature on ITE, Borko and colleagues (2007) identified four main genres of studies: (1) *effects of teacher education*, (2) *interpretive*, (3) *practitioner*, or (4) *design-based*. Research in the first two categories – *effects of teacher education* and *interpretive* – are more established and contributed significantly to the knowledge-base of ITE over the years. In contrast, *practitioner* research is a more recent addition to the field and includes action research, participatory research, self-study, and teacher research. This type of research is carried out by "by those who actually do the work of teacher education" (Zeichner, 1999, p. 8), namely teacher educators, who examine their own practice with the goal of understanding and improvement. The fourth genre – *design-based* research – emerged in the early 1990s and is rooted in the learning sciences. Design-based studies involve the development, implementation, and analysis of innovative practices or interventionist programmes for the purpose of theory-testing and building in teaching and learning (see Cobb *et al.*, 2003; diSessa and Cobb, 2004; Sandoval and Bell, 2004; Sandoval, 2014; Henrick, Cobb and Jackson, 2015).

The *effects of teacher education* research genre laid the foundations for research on ITE with a cause-and-effect orientation rooted in experimental psychological research and statistical analyses modelled on the natural sciences. As a result, studies of ITE effects in the 1960s and 1970s took on a "descriptive-correlational-experimental loop" (Borko, Liston and Whitcomb, 2007). In these process-product type studies, a teaching practice was described, variables correlated with pupil achievement identified, and interventions developed and then studied as teachers were trained to use them (Borko, Liston and Whitcomb, 2007). In later decades, the focus of such studies

broadened to include components of ITE and their effect on the development and practice of new teachers. Interest in this genre of research was revived with the spread of the standards and accountability movement, which focused policy and research attention on the *outcomes* of learning rather than in the inputs or process of learning. Experimental or quasi-experimental design using statistical analysis have often been used to study ITE effects because they attract policy attention (Cochran-Smith, 2004; Hostetler, 2005). However, the complex, contextual, and situational nature of teaching and learning has limited the generalizability of strictly quantitative studies and fuelled the expansion of interpretive research within ITE.

Interpretive studies of ITE, which emerged in the 1980s, attempt to uncover the local meanings by describing, analysing, and interpreting features of a specific situation while preserving its complexity. Interpretive researchers "attempt to capture local variation through fine-grained descriptions of settings and actions, and through interpretation of how actors make sense of their socio-cultural context and activities" (Borko, Liston and Whitcomb, 2007, p. 4). Interpretive studies may aim to address a variety of objectives, including (1) improving practice and/or programme design, (2) informing policy by highlighting influential or problematic contextual factors, or (3) contributing to theory-building. Studies with an interpretive lens have highlighted the significance of teachers' beliefs and knowledge in shaping their professional education experience and classroom practice. In addition, interpretive studies have shed light on our understanding of what learning occurs in ITE courses and field experiences, how teacher educators teach, and what is characteristic of high quality preparation programmes.

The focus on this review considers all these types of research in its goal of understanding how quality has been defined and measured in the field. However, in the past ten years, studies of ITE have most often focused on (1) issues of quality and the relevance of ITE, (2) what knowledge and skills trainee teachers should have, and (3) the role of teacher preparation in the ongoing professional learning of teachers (Clarke, Lodge and Shevlin, 2012). As a result, the literature reviewed here has mostly relied on effects of ITE research as well as interpretive studies of the impact of ITE. Interestingly, the widespread recognition of the value of high-quality teaching force has not, as Hanushek (2011) notes, "led to any consensus on the appropriate policies that should be followed to ensure that we have a good stock of teachers" (p.466). Thus,

attention has turned to ITE and how relevant and effective preparation programmes are in attracting and developing high-quality teachers. As a result, teacher educators and school personnel have been working to produce empirical evidence of practices and effects of ITE (Darling-Hammond, 2006a; Ludlow *et al.*, 2010).

2.6 The Impact of Governance on the Nature, Role, & Quality of ITE

In countries across the world, the state, universities, and the teaching profession play different roles and have varying levels of authority over ITE. As a result, the types of evaluations and judgements made regarding the quality of ITE in a given society must be understood in light of contextual particularities, including the governance structure, that shape them. Although many studies in the literature highlight the contextual particularities and politics of a given country's ITE sector (e.g., Shi and Englert, 2008; Caena and Margiotta, 2010; Furlong, Cochran-Smith and Brennan, 2011; Darling-Hammond and Lieberman, 2013), one comparative study – an examination of the governance structure of ITE in England and the Canadian provinces of Manitoba and British Columbia (Young, Hall and Clarke, 2007) – usefully illustrates how contexts shape the role and very nature of ITE. Drawing upon Dale's (1997) "dimensions" of governance (funding, regulation, and delivery) and Gideonse's (1993) "modes" of governance (political, institutional, or professional), Young and his colleagues analyse what impact particular governance arrangements – as semi-permanent "policy settlements" among competing interests in the field – have on the ITE sector. England, Manitoba, and British Columbia are all sub-regions of a larger country and have authority over their own ITE sector, and thus made for a fruitful comparative case-study of how governance settings affect each.

To do so, Young and his colleagues collected comparative data on three different dimensions of ITE governance: *funding* (i.e., amount, sources, and the basis and process of allocation), *regulations* (i.e., admissions criteria, certification, and accountability measures and processes), and *delivery* in ITE (i.e., site(s) of delivery, curriculum content and programme structure, and staffing recruitment, qualifications, and roles). Based upon the findings of these dimensions, the governance of each respective ITE sector was assessed to be predominantly politically-, institutionally-, or professionally-driven.

Young et al. (2007) determined that England's ITE sector was primarily politically governed, meaning that the state dominates teacher education policy. In this case, the central government not only claims legal jurisdiction over ITE but also exercises its authority through enabling structures (Ofsted and the TTA) rather than delegating such power to the university or a professional body. In contrast, ITE programmes in Manitoba are predominantly institutionally governed, meaning individual universities retain their own autonomy and determine the structure, curriculum, and degrees requirements within ITE programmes. In other words, control over ITE is located "within the university" and "its established structures of decision-making and accountability" (Young et al. 2007, p.85). Although Manitoba's provincial government, along with an advisory body drawn mainly from faculties of education, determine teacher certification parameters and regulates on elements of ITE (i.e., the length of the B.Ed. degree, minimum required time future teachers must spend in a school-based practicum; and the depth and breadth of subject area knowledge), the government cannot "effectively initiate substantial reforms of teacher education in the absence of university support" (Young et al, 2007, p.89).

Meanwhile, Young et al. (2007) characterized the ITE sector in British Columbia as predominantly governed by a self-regulating professional body – the British Columbia College of Teachers (BCCT). The BCCT, created in 1987, was given a mandate to set standards for teacher certification, professional conduct, discipline, and professional development. The body was also given the regulatory role of approving ITE programmes in the province. While this authority has, at times, been legally challenged by the universities seeking to protect their autonomy in designing ITE programmes and degree requirements, the members of the BCCT have continued to exercise their over-arching authority over entry into the profession and serve as a critical interface between the government, universities, and schools.

Young and his colleagues go on to suggest that, "the degree to which the state, the universities, or the profession dominate the governance of initial teacher preparation is directly related to different notions of teachers' work and the nature of the teaching profession" (Young et al. 2007, p.92). In other words, the form of governance in ITE has ramifications for how teaching, the teaching profession, and ITE are viewed, treated, and evaluated. Focusing on ITE, the authors link each form of governance to a particular "chararcterisation", or conceptualization, of ITE. Young et al. (2007) suggest

that, under *political governance* (e.g., England), ITE may be conceptualized as "a prescriptive practise designed to properly prepare new teachers to effectively implement a provincial or national agenda for schooling" and hence, an area in which "government control and supervision would logically prevail" (p.92). Under *institutional governance* (e.g., Manitoba), ITE may be treated as "a generative practice in which enquiry (and therefore knowledge production) is a defining feature of the teaching profession—a stance that would dovetail most easily with the mandate of the university" (p.92). Finally, under *professional governance* (e.g., British Columbia), ITE may be conceptualized as "a replicative practise of socialisation and induction – necessarily drawing on the embedded practical expertise of teachers" (p.92). Although Young et al. (2007) admit that these conceptualizations are not mutually exclusive and all three may be present in varying degrees, they argue that "the extent to which one dominates the other within teacher preparation has the potential to lead to significantly different notions of teacher education" (p.92).

     While these conclusions may be better understood and supported with more case-studies, the study highlights how competing interests and the governance structure in a given jurisdiction shape the nature and role – and by extension, the evaluation and quality perceptions – of ITE. The wider purpose, role, and governance structure of an ITE system shape the set objectives, measuring tools, and judgments of ITE quality. This is an important point to keep in mind as the literature review highlights different models of quality evaluation used in various contexts. In addition, the importance of context as a powerful lens through which quality must be judged also requires states to consider their own goals, philosophy, and governance structure when determining their own indicators and frameworks with which to measures quality in ITE.

3. Current Approaches to Conceptualizing & Measuring ITE Quality

3.1 Notable Challenges of Evaluating Quality in ITE

Efforts to measure the quality, or effectiveness, of ITE in the past have faced many difficulties. The first is the lack of a shared research base, common theoretical lens, or knowledge base with which to approach the study of ITE. Blömeke et al. (2008) summarise the key difficulties in developing systematic studies of the effectiveness of ITE:

> There are… only a few empirical studies and these are mostly small-scale studies or analysis of policy documents… research is often short term, of non-cumulative nature, and conducted within the own training institution…. What matters the most is probably that teacher education research lacks a common theoretical basis, which prevents a convincing development of instruments and makes it difficult to connect studies to each other.
>
> (Blömeke *et al.*, 2008, p. 719)

In a more recent review, Sleeter (2014) examined 196 articles published in 2012 in four leading international teacher education journals. She reported that she "did not see evidence of an emerging, shared research program designed to inform policy" (p.151). Sounding similar to Blömeke and colleagues, she argues:

> The problem [...] is that the weight of the research, being fragmented, often narrowly focused, and usually not directly connected to a shared research agenda on teacher education, does not position teacher educators strongly to craft an evidence-based narrative about teacher education that might counter policies and reports like the NCTQ's [National Council on Teacher Quality, Washington D.C. think tank].
>
> (Sleeter, 2014, p.152)

As a result, Sleeter (2014) advocates for (1) collaborative studies of ITE that bring together a number of researchers, (2) studies that are embedded and connected to a wider, shared research agenda in teacher education and that will contribute to building a sound knowledge base for the field, and (3) studies which produce useful evidence and conclusions of interest to policy-makers responsible for shaping education.

Another common challenge to evaluating programme quality is how to conceptualise, operationalise, and measure outcomes in ITE. Still, teacher educators continue to develop tools and strategies to assess the impact of their efforts in ways that

take into account the complexity of teaching and learning and provide a variety of perspectives on the process of learning how to teach. Drawing upon the work of Cochran-Smith (2001), Darling-Hammond (2006a) outlines three ways in which the outcomes of teacher education may be measured, or evaluated – through evidence about teacher test scores (e.g., tests related to teachers' knowledge), evidence of the professional performance of teacher candidates, or evidence of the impacts of ITE on teaching practice and student learning in the classroom. Commonly used measures to gauge teachers' abilities as well as the effectiveness of teacher educators themselves include: (1) candidate performance in courses, (2) candidate performance in student teaching, (3) various assessments used within programmes, (4) data on entry to ITE programmes, (5) retention of candidates in teaching, and (6) candidates' perceptions of their preparedness as well as their employers' perceptions of their effectiveness once they are teaching (Darling-Hammond, 2006a). A handful of programmes have attempted to measure teachers' "impact" in the classroom through an analysis of pupil learning based on changes in student attitudes or behaviour, work samples, performance assessments, or student scores on standardized tests. These ways to measure the knowledge and performance of teacher candidates and teacher educators illustrate the difficulty in capturing and measuring a complex and situated practice such as learning to teach.

Finally, another challenge in assessing ITE quality lies in the collection, interpretation, and management of data. Evidence of the impact, or "effectiveness", of ITE is difficult to collect and interpret for at least three reasons (Darling-Hammond, 2006). First, it is difficult to develop or obtain "comparable pre-measures and post-measures of student learning that can gauge change in valid ways that educators feel appropriately reflect genuine learning" (p.121). Second, while finding valid measures is challenging, it is also difficult to link changes in student attitudes or performance to an individual teacher given the variety of other factors influencing children. Lastly, it is a challenge to attribute the knowledge and/or practices of teacher candidates or graduates directly to ITE programmes since outside factors such as school context, peers, and past experiences shape teachers' beliefs and practices. Despite these challenges in collecting and analysing data, such evidence is valuable for assessing the quality and impact of ITE. Thus, storing, managing, and sharing data becomes equally important to benefit other ITE programmes, educational researchers, policy-makers, and other stakeholders.

Therefore, while the challenges in assessing the quality and impact of ITE programmes are many, the value and need to do so makes the endeavour all the more necessary. In recent years, a number of studies have aimed to overcome such obstacles to provide valid evidence of the role and impact of ITE in producing quality teachers. In the next section, the components deemed to contribute to ITE quality are reviewed, and the evidence commonly collected to measure them is discussed.

3.2 Commonly Assessed Components Contributing to ITE Quality

To examining how to measure ITE quality in the Scottish context, it is helpful to begin by reviewing and discussing the components examined and the measures most commonly utilized in previous studies on the topic. This will provide a general overview of the literature – including genres and common measurements of ITE quality – and, in so doing, provide a solid base from which to examine more complex and innovative frameworks for measuring ITE quality. To begin, in their comprehensive report on teacher preparation evaluation, Feuer and colleagues (2013) outline six attributes related to ITE quality and the commonly used evidence to measure them. These six attributes include (1) *admissions and recruitment criteria*, (2) *quality and substance of instruction*, (3) *quality of student teaching experience*, (4) *faculty qualifications*, (5) *effectiveness in preparing new teachers who are employable and stay in the field*, and (6) *success in preparing high-quality teachers*. These areas and the evidence used to measure them represent both inputs and outcomes of ITE programmes as well as attempts to capture the design, content, and process of learning how to teach through ITE. Each of these areas, and the evidence used to account for them, have both strengths and weaknesses.

First, the *admissions and recruitment criteria* ITE programmes use is typically measured through academic results or exam scores of incoming cohorts as well as data on the percentage of minority candidates recruited and percentage of candidates recruited for specialised or shortage subject areas. *Admissions and recruitment criteria* are often associated or linked with the *selectivity* of ITE programmes. Selectivity is commonly conceptualised in terms of the percentage of applicants who applied that are admitted into the programme. In addition to such admittance rates, the overall quality of the applicant intake is also a mark of selectivity. Incoming cohorts that are high-

achieving academically are considered "high quality" recruits. An ITE programme's degree of selectivity is also treated in studies as a measure that reflects the quality and/or rigor of an ITE programme. This link between selective recruitment and ITE outcomes is supported by research that suggests that more selective programmes may produce more effective teachers (Kane, Rockoff and Staiger, 2008; Henry, Bastian and Smith, 2012). However, according to Kane (2008), this potential correlation only accounts for a small amount of variation in teacher effectiveness and is often influenced by several factors beyond the control of ITE programmes (e.g. institutional location, prestige, or resources, local status of the teaching profession, labour market needs, etc.) (Labaree, 2010). On this topic, it is worth noting that there is evidence that high-quality teacher candidates, often referred to as the "best and the brightest", may leave the profession in higher numbers than other recruits (Boyd *et al.*, 2006; Xu, Hannaway and Taylor, 2007; Kelly and Northrop, 2015).

Secondly, the *quality and substance of instruction* of ITE programmes is frequently measured through analysis of course syllabi. This measure can help identify inconsistencies, redundancies, or gaps in the coursework as well as its apparent connection or integration with fieldwork. However, this measurement is limited in that it may not accurately reflect what is actually taught in class and assessing the quality or efficacy of such syllabi may be difficult given debates regarding what knowledge base teachers need to be effective in the classroom. Other evidence often measured (besides course syllabi) used to measure *quality and substance of instruction* includes the review of other materials (e.g., textbooks, assignments) and other programme documents related to course offerings and hours. Such document reviews use pre-existing data to provide descriptive details about the design and processes of ITE programmes. Document reviews are often a faster and less costly way to gain programmatic information compared with observing lectures and the delivery of other programme components. However, such data may fail to capture what actually occurs in practice (rather than what is planned or intended) and requires the development of analysis criteria and the carrying out of the time-consuming task of assessing numerous documents.

Thirdly, the quality of student teaching experience of ITE programmes is typically measured in four ways: (1) fieldwork policies (e.g., required hours, supervision policy, observation frequency, etc.), (2) qualifications of placement school mentors, (3) observations of student teaching, and (4) surveys of teacher candidates regarding their

fieldwork. The first two measures, fieldwork policies and mentor qualifications, provide a basis on which to compare programmes and judge design. However, these measures likely do not capture the impact of the student teaching practicum or quality of mentoring provided by cooperating teachers. The latter two measures – observations of student teaching and surveys of teacher candidates regarding their fieldwork – do provide more qualitative insight into the effectiveness of the practicum in ITE. Observations of practice are most useful when the evaluators are well-trained and share a commonly designed and agreed upon rubric for identifying and recording what they witness (Darling-Hammond, 2006a). Meanwhile, surveys of ITE graduates regarding their experience of the teaching practicum may be particularly helpful as they provide a closer look at the quality of student teaching experience offered. However, while such surveys represent a relatively easy and inexpensive way to capture feedback, the results of survey are limited measurements as they rely on perceptions rather than objective measurement. In addition, this form of measurement requires a timely administration and high response rate.

Fourthly, measures of ITE faculty qualifications typically include the percentage of faculty with advanced degrees and the percentage of faculty that are full-time, part-time, or adjunct/supplementary. The fact that these are the only two measures mentioned by Feuer and colleagues (2013) highlight the neglect of teacher educators as a factor in ITE quality. Teacher educators, particularly as an attribute of quality ITE programmes, are much less frequently examined, if at all, in the current literature on quality ITE frameworks (Snoek, Swennen and van der Klink, 2010). The two types of evidence named above are simple measures that do not begin to capture the variation, professionalism, or effectiveness among teacher educators. Neglect of this area of ITE is an ironic trend given the global consensus that quality teachers, and by logic extension teacher educators, matter (Goodwin et al., 2014). This point will be returned to in a discussion further below on the gaps, or missing links, in current ITE frameworks.

Fifthly, the extent to which ITE programmes are *effective in preparing new teachers who are employable and stay in the field* is commonly measured by pass rates on licensure/qualification exams as well as data on the hiring and retention rates of ITE graduates. Measures of licensure/qualification exams vary from subject knowledge exams (e.g., PRAXIS tests in the U.S.) to more competency-based assessments (e.g., qualified teacher status awards in England, or Standard for Full Registration in

Scotland), and thus are more context-specific than data on hiring and retention of ITE graduates. Data on employment outcomes is typically measured by graduate placement rate and retention rate for a graduate's first three years. Such measures have been linked by some studies to ITE programme duration and type (e.g., Darling-Hammond, 2000; Ingersoll, Merrill and May, 2012) as well as teachers' satisfaction with their preparation programme (DeAngelis, Wall and Che, 2013). However, such measures often do not take into account conditions in the teacher labour market, personal life circumstances and employment preferences of new teachers, or how school resources and contexts affect employment outcomes of ITE graduates (Kumashiro, 2015). Tatto et al. (2016) also notes that programmes that most likely will show positive employment outcomes are those with more resources to invest in such longitudinal efforts, have stronger social and/or institutional networks through which to place and support new teachers, and those programmes that have a guarantee of employment such as Teach For America (the U.S. version of Teach First) and other work-based routes into teaching.

Finally, ITE programmes' *success in preparing high-quality teachers* is commonly measured by teacher performance assessments administered toward the end of the programme or in the year(s) after graduation, through graduate and/or employer surveys, and sometimes through value-added approaches to measuring teachers' impact on student learning. Teacher performance assessments of teacher candidates and/or graduates, which take a variety of forms depending on programme ethos and context, are becoming more popular as a way to capture the complex and situated nature of teaching. Hence, they are often considered as more authentic measures of the outcomes of ITE. The Performance Assessment for California Teachers (PACT) is a prime example of a teacher performance assessment developed and shared across one state's ITE sector (Falk, 2013; Merino and Pecheone, 2013; Wilkerson, 2015). Other performance assessments have been developed and spread by professional associations and are increasingly used to judge ITE quality (Drew, 2016). However, the drawbacks of such assessments include the significant time and effort from the teacher candidates and evaluators in producing and evaluating them as well as the risk of being subjective if not systematically implemented (Darling-Hammond, 2006a; Coggshall, Bivona and Reschly, 2012).

In addition to teacher performance assessments, evidence of *high-quality teachers and teaching* among recent ITE graduates is also frequently measured by

ratings of graduates' performance by employers (e.g., principals) or, as commonly done in the U.S., through valued-added scores which aim to measure effectiveness of teachers based on students' standardised test scores. An additional tool of measuring ITE quality that may be placed in this category includes surveys of ITE graduates regarding their programme experience, their current practice, and their feeling of preparedness and efficacy among other things (e.g., Darling-Hammond, 2006a; Ludlow *et al.*, 2008; Clarke, Lodge and Shevlin, 2012). When combined, surveys of ITE graduates and ratings from their hiring principals provide two types of feedback of the perceptions of new teachers' knowledge and teaching ability. However, Tatto et al. (2016) argue the record of using surveys of graduates and employers rests on the problematic assumption that higher levels of satisfaction imply higher levels of ITE quality. Tatto et al. also point out that graduates' level of satisfaction with their ITE may be linked to the rigor of the coursework and/or graduation requirements. Thus, they highlight the need to combine measurements of ITE graduates' satisfaction with evidence of their levels of knowledge attained to gain a clearer picture of graduates' perceptions and learning.

A number of studies have investigated whether employer surveys have accurately assessed the effectiveness of new teachers, and the evidence is mixed. While Coggshall, Bivona, and Reschly (2012) found strong correlations between principals' assessments of their teachers and teachers' value-added measurement scores, economists Jacob and Lefgren (2008) employed another quantitative approach to examine how accurately principals can distinguish between more and less effective teachers. By analysing demographic variables for students, teacher characteristics (i.e. age, experience, educational background, etc), and survey responses from principals, Jacob and Lefgren found that principals were most able to identify high- and low-performing teachers (teachers whose students had the largest and smallest standardized achievement gains). However, principals' accuracy in judging teaching ability was less when assessing teachers whose students showed more mediums gains as measured by value-added measurement (VAM) (and hence were not in the extremes of the distribution of gains). Although the study has its limitations, it highlights the potential problems of relying on principals' surveys to reliably judge ITE graduates' knowledge and ability.

In a related study examining the accuracy of judging teachers through observation, Strong, Gargani and Hacifazlioğlu (2011) conducted three experiments designed to determine (1) whether expert evaluators can observe and correctly identify

highly effective teachers and (2) on what basis they make their judgments. The purpose of their study was to understand under what conditions and with what criteria judges' evaluations are most accurate. While teacher evaluators in all three experiments produced consistently similar ratings of observed teachers' effectiveness, those judgments revealed an inability to identify accurately 'effective' teachers. However, certain criteria related to teachers' instructional behavior *did* reliably enable the identification of effective teachers.

As a result, Strong and his colleagues conclude that even experienced evaluators are not able to identify, with accuracy, "effective" teachers as measured by test scores – an indicator the authors admit is limited but important to a wide variety of stakeholders. They suggest this is due to psychological and cognitive influences affecting human behaviour and judgment. These influences include (1) *confirmation bias*, which is the tendency for individuals to notice and pay more attention to evidence and experiences that confirm, rather than challenge, their pre-existing beliefs or notions, (2) *motivated reasoning*, which is the tendency for individuals to look more skeptically at evidence that do not match our already-held beliefs, and (3) *inattentional blindness*, which is the tendency for individuals engaged in an attention-demanding task to *not* notice certain evidence or occurrences they would otherwise. Finally, the researchers also conclude that it is both "desirable and possible to develop a new measure that does produce accurate predictions of a teacher's ability to raise student achievement test scores" (Strong, Gargani and Hacifazlioğlu 2011, p.367).

This led to Gargani and Strong (2014) to attempt to develop teacher evaluations that would be more faster, cheaper, and more accurate than previously used frameworks. They developed and tested a teacher observation method – the Rapid Assessment of Teacher Effectiveness (RATE) – that uses a six-item rubric and requires only 4 hours of evaluator training. In experimental studies testing this tool, teacher evaluators were able to identify effective teachers after only twenty minutes of lesson observation with an accuracy rate of assessment higher or the same as others using more complicated and time-consuming evaluation frameworks. However, our ability to develop and use evaluation instruments that identify 'effective' teachers are, of course, only one (narrow) measure of teacher quality.

Regardless of debates regarding the accuracy of human judgments of teaching ability, both types of surveys – those of ITE graduates and their hiring principals – are

common. In their extensive review of ITE evaluation measures, Feuer et al. (2013) reported that, when asked about their views and experiences of their ITE programme, most graduates may provide some insights but these are ultimately of limited use in determining ITE quality due to participants' subjectivity and selectivity bias. Moreover, Darling-Hammond (2006) evaluated the newly redesigned Stanford Teacher Education Program (STEP) by collecting and analysing data over a five-year period using a range of measures, including surveys of STEP graduates and the principals that hired them. Notably, survey results indicated that principals overall were highly positive about STEP graduates while the graduates themselves were more modest in their self-assessment. This outcome again highlights the subjective and potentially biased nature of survey data. This does not negate its value in assessing teachers' effectiveness as part of ITE quality but does emphasize the need to collect and combine different types of evidence to measure this outcome. In other words, the conclusion regarding these surveys is that they can be useful when triangulated with other types of data but alone are not reliable or conclusive as indicators of ITE quality.

The final measure of ITE programmes' *success in preparing high-quality teachers* is value-added approaches, increasingly commonplace in the U.S., which attempt to estimate a teacher's contribution to students' progress over time. Value-added measures (VAM), or modelling, link teachers to students' standardized tests scores, and their teaching effectiveness is measured through a statistical regression model. The regression model controls for additional factors (i.e., prior student achievement and student, family, and neighbourhood characteristics) to determine how much value-added learning a given teacher produces compared to a calculated "average" teacher. By accounting for additional variables outside teachers' control, the regression models "statistically isolate the contribution of programs, and teachers to *growth in student achievement* at a given grade level from all other sources of student achievement growth" (Muñoz and Chang, 2007, p. 147 emphasis in the original). Proponents of VAM argue that value-added models provide a common tool with which compare the effects of different programmes, enabling researchers to quantitatively identify differences between graduates (Gansle, Noell and Burns, 2012). More controversially, however, it has been argued that they result in a disproportionately negative impact on pupils from disadvantaged and minority backgrounds (Mangiante, 2011) and emphasise

learning that can be measured quantitatively, to the detriment of other aspects of education (Biesta, 2009).

VAM approaches have also been criticised as imperfect and hence unreliable tools for data collection on the bases of attribution error and bias. First, VAM data is difficult to calculate. There is not one commonly agreed upon approach to calculating VAM (Guarino *et al.*, 2015; Walsh and Isenberg, 2015) and only a few states have "fully-tested and functional" data systems required to collect and analyse for VA (Data Quality Campaign 2012). Secondly, only a small sub-set of ITE graduates can be assessed through VAM as the method relies on standardized tests with are only given in certain subject areas and at only a few grade levels. Thirdly, VAM measures are limited in that they do not provide any guidance or indicators on *how* to improve programmes. Another concern regarding the usefulness and reliability of VAM centre on the fact that ITE graduates are typically not assigned students at random (the least experienced teachers are often assigned to teach lower performing pupils) (Rockoff and Speroni, 2010). Finally, critics of VAM argue that the model does not adequately account for other "desired student learning outcomes" beyond tested knowledge. Interestingly, recent studies employing VAM approaches have shown little variation in graduates across different teacher training programme and greater variation among graduates of the same programme (Mihaly *et al.*, 2012; Goldhaber, Liddle and Theobald, 2013; Koedel *et al.*, 2015).

3.3 Evaluating ITE Quality: Competing Research Orientations

The review of components and measures commonly used to assess ITE quality provides the foundation from which more complex studies of ITE quality have been built. Studies of ITE quality can also employ one or more theoretical approaches. Blömeke et al. (2008) highlights the common *educational-sociological perspective*, reflected in AERA panel research handbook edited by Zeichner and Cochran-Smith (2005), which focuses on characteristics of teacher education programmes and investigates their effects. In contrast, Blömeke et al. (2008) take an *educational-psychological perspective* "by focusing on professional competence of teachers and examining influences on this" (p.719). As a result, Blömeke and his colleagues take the teacher candidates as their focus – measuring the cognitive components of professional

competence such as types of professional and latent knowledge, personal beliefs and personality traits, and the development of professional skills. In this way, Blömeke and colleagues conceptualize the central criterion of "effective teacher education" as "professional competence of future teachers" (p.719), and thus track the development of student teachers' professional competence in teaching through their ITE programme and beyond. More elaborate quality assessment frameworks can take on both a sociological as well as psychological orientation by developing studies and measures that track both the effects of ITE components as well as candidates' development of competence.

At this point, it is also worth noting that studies of ITE quality can and do take a variety of forms. In a review of 196 articles featured in four leading international journals of teacher education in 2012, Sleeter (2014) found that 86%, reported on research that used the following designs: survey research; qualitative case studies; analysis of documents, videos, or discourse; small-scale experimental or quasi-experimental research; Q-sort methodology; and mixed-methods research. Thus, the literature on how ITE quality has been measured is wide and varied in its orientations, purpose, focus, and sources of evidence. With this awareness, this review now moves on to examining important trends and contrasting approaches and frameworks in this literature.

4. Profiles of ITE Quality Frameworks and Studies

In this section, profiles of different frameworks for measuring, or evaluating, the quality of ITE are reviewed and discussed. These profiles, or case studies, are attempts to capture and discern what *is* and *is not* working within ITE programmes or, alternatively, which types of programmes are more effective in creating high quality teachers. To understand the context and methodology of these case studies as well as benefit from the lessons that may be gleaned from them, each section is rather detailed in describing the frameworks for assessment developed. These represent a variety of contexts and approaches and are only a sample of what examples are out there. These particular examples were chosen for their unique and contrasting approaches, which highlight the many different ways "quality" may be conceptualised and measured.

4.1 "Powerful" ITE Quality Indicators: Framework Derived from Ideal Models

In considering ways to enhance the quality of ITE programmes, some researchers have started with the question: *What is quality in ITE?* and more specifically, *What do high-quality ITE programmes look like?* While this answer will certainly vary by national context, one particular study by Darling-Hammond (2006c) attempted to answer the question through comparing a diversity of highly effective ITE programmes in the U.S. This study has led to a strand of research into features, and hence possible measures, of "powerful" ITE programmes. In her study, Darling-Hammond (2006c) examined the goals, strategies, content, and processes as well as the organizational features, resources, and relationships of "powerful" teacher education programmes at seven higher education institutions[1]. The sample represented a diverse set of institutions, that are public and private, undergraduate and graduate, large and small, but have one thing in common. These programmes, according to Darling-Hammond's data, produced new teachers who are "extraordinarily" well prepared to become effective teachers from their first days in the classroom. Darling-Hammond argues that despite outward differences, these programmes shared seven common features:

---

[1] The seven universities whose ITE programmes were in the study were (1) Alverno College in Milwaukee, Wisconsin, (2) University of California at Berkley, (3) University of Virginia in Charlottesville, (4) the University of Southern Main, (5) Wheelock College in Boston, (6) Bank Street College in New York City, and (7) Trinity University in San Antonio, Texas.

1. a common, clear vision of good teaching that permeates all course work and clinical experiences, creating a coherent set of learning experiences;

2. well-defined standards of professional practice and performance that are used to guide and evaluate course work and clinical work;

3. a strong core curriculum taught in the context of practice and grounded in knowledge of child and adolescent development and learning, an understanding of social and cultural contexts, curriculum, assessment, and subject matter pedagogy;

4. extended clinical experiences—at least 30 weeks of supervised practicum and student teaching opportunities in each program—that are carefully chosen to support the ideas presented in simultaneous, closely interwoven course work;

5. extensive use of case methods, teacher research, performance assessments, and portfolio evaluation that apply learning to real problems of practice;

6. explicit strategies to help students to confront their own deep-seated beliefs and assumptions about learning and students and to learn about the experiences of people different from themselves;

7. strong relationships, common knowledge, and shared beliefs among school- and university-based faculty jointly engaged in transforming teaching, schooling, and teacher education

(Darling-Hammond, 2006b, pp. 305–306)

In addition, Darling-Hammond highlights the need for coherence and tight integration between coursework and clinical practice as well as extensive and intense supervision and modelling from expert teacher mentors who have experience serving diverse students (2006b, 2006c). Moreover, Darling-Hammond argues it is equally important to ensure universities have cultivated a strong, supportive, and collaborative relationship – geared toward learning and improving practice in-sync with the ITE programme – with the schools in which they place candidates. Professional development schools, lab schools, and school reform networks are simply some initiatives she mentions that have, at times, achieved this ideal.

Darling-Hammond's study, involving a number of education researchers, led to a widening conversation and research agenda focused on identifying key features of "powerful" teacher education. In this context, the term "powerful" evolved and is now

used to refer a teacher education programme that achieves a number of goals, it: (1) "prepares teachers to teach students from diverse racial, ethnic and linguistic backgrounds", (2) "counters lay theories and beliefs about teaching", (3) "is in line with views of teaching as a profession that can be learned and developed over time", (4) "prepares teachers learning in practice with a deep theoretical and research base", and (5) "can prepare teachers who remain in teaching longer than their peers" (Hammerness and Klette, 2015, p. 5). Although the term "powerful" was initially used by Darling-Hammond (2006c) in her study of exemplary ITE programmes, it is also useful in describing the quality framework that evolved from it and related studies.

Building from Darling-Hammond's study and other research in the U.S. (e.g., Darling-Hammond, 2006a; Grossman *et al.*, 2008) and the Netherlands (e.g., Brouwer and Korthagen, 2005), Hammerness (2013) narrowed down the effective features of successful ITE programmes and argues that powerful teacher education programmes share three key features: (1) the promotion of *a clear vision* of teachers and teaching, (2) programme *coherence*, both conceptually and structurally, "reflecting a shared understanding of teaching and learning among faculty and students" (p. 400), and (3) a strong core curriculum that is closely linked to opportunities for learning that are grounded in teaching practice.

With this three-point framework, Hammerness (2013) examined nine ITE programmes at six different higher education institutions in Norway. She interviewed programme leaders or department heads and at least one other faculty member, asking about the espoused programme purpose and goals for future teachers and their practice, the coherence of programme to item, and links between coursework and clinical practice. She also examined course syllabi and materials for evidence of the programme vision and links between theory and practice. Hammerness (2013) found that there were three themes that emerged and reflected Norwegians' view of teachers as subject matter experts, classroom leaders, and ideally, researchers. Yet, these three core ideas served more as common reference points for the individual visions of the teacher educators interviewed rather than a deliberate, articulated, or shared programme vision. In fact, some interviewees noted the difference in views on teaching among faculty members, particularly between subject matter and pedagogy specialists. Finally, Hammerness found that programmes generally lacked coherence and meaningful opportunities for teacher candidates to learn in the context of practice were limited or absent.

While Hammerness (2013) examined the programme vision and coherence in ITE in one country, Canrinus, Bergem, Klette, and Hammerness (2017) focused on measuring the *coherence* within ITE programmes in five different countries from a student teacher perspective. Canrinus et al. (2017) define *programme coherence* as "a process, in which all courses within a programme, whether theoretical or practical, are aligned based on a clear vision of good teaching" (p. 315) and argue that "a key means of determining coherence is student perception" (p. 316). The researchers stress that coherence is not an end product but a process that requires continuous adaptation and fine-tuning of courses, placements, and ongoing conversations among ITE faculty regarding the vision of good teaching that should be underpinning the entire programme. Canrinus et al. (2017) argue that programme coherence is important for helping teacher candidates develop a strong sense of professional identity (Rogers, 2011) and that, while programme designers and faculty may believe a programme reflects coherence, candidates' perceptions may differ (Clift and Brady, 2005).

To assess the perception of *programme coherence* among student teachers, Canrinus et al. surveyed 486 student teachers in highly-regarded university-based ITE programmes in five different contexts – Oslo (Norway), Stanford (USA), Helsinki (Finland), Santiago (Chile), and Havana (Cuba). Through these surveys, Canrinus et al. (2017) aimed to discover: (1) To what extent do student teachers experience their teacher education programme as coherent? (2) What similarities and differences between the programmes regarding their students' perception of coherence in the respective programmes can be observed*?* (p. 316). The surveys used were based, in part, on surveys used by Boyd et al. (2006) in their New York City Pathway Study to study the varying characteristics and learning opportunities in different types of teacher preparation programmes. The survey developed and used by Canrinus et al. (2017) contained 19 items that were rated on a four-point Likert scale and assessed the coherence across courses, between courses and fieldwork (teaching practicum), and student teachers' opportunity to connect various parts of the programme.

The researchers found both similarities and difference across programmes but discovered student teachers in the Stanford programme reported significantly higher ratings of *programme coherence* overall. Stanford student teachers, along with student teachers in Havana, reported the coherence between courses to be strong while student teachers in Norway reported the coherence between courses to be weak. In addition,

Stanford student teachers indicated that they experienced greater coherence between their coursework and field experiences compared to all other student teachers. Meanwhile, student teachers in Santiago reported the least coherence between their courses and field placement. Lastly, Stanford student teachers indicated they experienced significantly more opportunities to connect various parts of the programme while those in Helsinki reported having the fewest opportunities to do so (though not significantly different than student teachers in Havanna, Santiago, and Oslo).

Overall, the findings of the surveys indicated that student teachers experienced a reasonable amount of coherence in their programmes with coherence between courses more evident than between courses and fieldwork. Canrinus and her colleagues then discussed the particularities of each context that likely led to the trends identified in the survey outcomes. Notably, they revealed that the Stanford ITE faculty had invested considerable time and resources redesigning and evaluating its programme over the past 15 years, a case that this review will examine in more detail later on. The findings of the study also highlighted the enduring challenge for ITE progarammes in bridging the theory-practice gap between courses and fieldwork, which affects the coherence of the programme as a whole.

Feiman-Nemser and her colleagues (2014) built on these principles of vision-focused, coherent ITE programmes that provide well-integrated teaching practice by examining a set of mission-driven schools to understand how effective teaching is inspired through ITE. The researchers draw on the *Choosing to Teach* project, a detailed multi-year study of three selective, mission-driven teacher preparation programmes – the Urban Teacher Education Program at the University of Chicago, the Alliance for Catholic Education at the University of Notre Dame, and the Jewish-oriented Day School Leadership through Teaching programme at Brandeis University. This study traces each programme's impact on graduates during their first few years of teaching by tracking graduates' professional growth and how they face challenges in their respective school sectors. Notably, in this study, Hammerness (2014, p. 104) identified three ways in which ITE faculty expressed visions – as shared expressions of their personal and programmatic commitments. These three visions were described as: (1) *teaching as service*, which conceives of teaching as only one of many opportunities to "give back" or contribute to society, (2) *teaching as social justice*, which conceives of teaching as a direct means of addressing social inequities, (3) *teaching as a practice*,

which focuses on teaching as a profession with a knowledge base and set of practices that can be learned and developed over time. She found these distinct visions for teaching at times, overlapped, but still captured differing visions as expressed by programme faculty.

These studies of the *vision*, *coherence*, and *coursework integration with meaningful teaching practice opportunities* that are characteristic of exemplary preparation programmes represent one conversation, or strand of research, that has emerged in the field since the mid-noughties. As this research has accumulated, Hammerness and Klette (2015) describe this particular three-point framework as "a set of powerful, well-researched indicators" of ITE programme quality that can be applied to any type of teacher preparation programme, of any size, and in any context. The conversation and research focused on identifying key features, or measureable indicators, of "powerful" teacher education and ways to assess them continues to grow. Proponents of this "powerful" ITE framework of quality indicators argue it reflects how new teachers learn most effectively to become teachers, and references Stanford's ITE programme among others as a prime example. The framework also benefits from an emerging empirical base that provides support for the value of these core features and has begun to inform teacher educators as to how they might improve the quality of their ITE programmes.

Still, this framework is limited as it is rooted in a general consensus, or theory, of what features make an ITE programme effective. It provides guidance by identifying what *should* be in place and happening in ITE and offers examples of how to measure such. Meanwhile, others have come up with alternative ways of conceptualizing ITE quality and measuring the impact of programmes in more contextually- and programme-specific ways, which are discussed below. Nevertheless, the seven programme characteristics of effective ITE programmes named by Darling-Hammond (2006c) and the three-point framework proposed by Hammerness and colleagues (2016) may provide teacher educators with a starting point in discussions of what quality in ITE looks like for their programme as set in local contexts.

4.2 Comparing Quality Evaluation Approaches in ITE from the U.S.

It is worth examining the way in which ITE has and is being assessed currently in the U.S. for two reasons. First, there are more than 1,300 teaching training institutes across the country that adhere to different state standards and can be accredited by more than one national body. In addition, numerous alternative pathways into teaching have been developed in most states over the past 30 years. Thus, the volume of research and number of approaches to evaluating the impact of ITE is relatively higher compared to other countries, and thus the context offers lessons and insights into how others go about studying the quality of ITE. Secondly, although the U.S. is unique for the frequency in which statistical value-added models (VAM) are used to judge the effectiveness of teachers, as well as the effects of ITE programems and alternative pathways, there are lessons offered by such studies that even those critical of VAM may consider useful. Therefore, in the following discussion, the context of evaluation studies of ITE is briefly considered and then examples of contrasting frameworks and tools are examined in more detail. Lessons from these contrasts are then drawn out and summarized.

First, it is worth describing, briefly, the context in which ITE evaluation is occurring in the USA. In the U.S., where authority and governance of public education is purposely fragmented, a number of approaches to evaluating ITE programmes have evolved in the last ten years. Feuer et al. (2013) categorize such approaches based upon the bodies responsible for carrying them out: (1) the federal government, (2) national accreditation bodies, (3) state governments, (4) independent organizations and media outlets, and (5) ITE programmes themselves. Feuer and colleagues (2013) go on to point out that many of these institutions and organizations use "multiple approaches and data methodologies" yet "actually rely on similar sources of data and on similar methods for converting data into judgments of program quality" (p.23). In addition, all these types of evaluations have one or more of three main (and, at times, overlapping) purposes: (1) ensuring accountability and monitoring, (2) providing consumer information, and (3) improving teacher preparation programs. These points serve to highlight the importance of considering the *purpose, audience*, and potential *agenda* of any evaluation or effectiveness study in ITE.

In order to draw lessons , trends in *state-led, policy-oriented* ITE evaluations in various U.S. states are outlined (Coggshall, Bivona and Reschly, 2012) and then compared with two different *practitioner-developed, improvement-oriented* evaluation

frameworks used in ITE programmes at Boston College (Enterline *et al.*, 2008; Ludlow *et al.*, 2008, 2010) and at Stanford University (Darling-Hammond, 2006a).

4.2.1 State-led, Policy-oriented Evaluations for ITE Accountability in the U.S.

In the early 2000s, the passage of *No Child Left Behind* rapidly spread and greatly increased the use of standardised testing to measure student outcomes. This development, combined with the Obama administration's *Race to the Top* initiative and a focus on improving teacher preparation, encouraged states to create quality evaluation frameworks for ITE programmes and their graduates. In their detailed policy brief, Coggshall et al. (2012) profile six U.S. states – Louisiana, Texas, Tennessee, North Carolina, Ohio, and Florida – that they argue "have led the way in changing how they evaluate the effectiveness of teacher preparation programs" by "combining measures to construct a more complete picture of the quality of teacher preparation programs" (p. 23). All use VAM approaches in some way or another as stipulated or encouraged by state policy regulations. However, the variety of ways in which states combine this data with other measures is intriguing.

Louisiana, which was the first state to use VAM in its assessment of ITE, compared new teachers' VAM data with ITE output data such as graduate surveys and knowledge tests to measure ITE effectiveness. Texas took also took VAM scores of ITE graduates and combined them with data from certification examination pass rates, evaluations of graduates' performance, and data on the programmes' field supervision (frequency, duration, and quality) to arrive at judgments of ITE quality. Similarly, Tennessee combined VAM data on graduates' placement and retention rates and subject knowledge test scores to determine which ITE programmes produce effective teachers.

In (rather slight) contrast, North Carolina relied on VAM scores exclusively (though using two types of statistical models) to compare effectiveness across different types of ITE programmes, and Ohio brought together policy-makers, private corporations, and teacher educators from across 51 higher education institutions to create the most comprehensive ITE evaluation of the six profiled states. The state-wide framework developed in Ohio measured ITE quality using VAM scores combined with the results of surveys, performance assessments, and subject knowledge tests of graduate teachers as well as appraisals of ITE programmes and feedback from schools. This

qualitative and quantitative data informed five separate studies on pre-service teachers, novice teachers, alternative licensure, experienced teachers along with a structural equation modelling study (Lasley, Siedentop and Yinger, 2006). Based on the analysis and synthesis of this data, Ohio designed report cards for higher education. Florida, meanwhile, was still in the process of determining its measurements for ITE quality but was profiled by Coggshall and colleagues (2012) because of its extensive recruitment of stakeholders and expertise to inform the design of its new accountability system and performance targets for ITE.

These efforts and resulting frameworks, developed at the state-level with the aim of informing (and, to some extent, enforcing) policy, represent an instrumental approach to determining ITE quality. While each state draws on multiple data sources and methods, the evaluation outcomes provide snapshots of effectiveness and efficiency across a wide range of programmes. These types of frameworks, however, are not well-designed to inform deeper insights into why and how ITE programmes are effective or not. However, quality frameworks and studies developed by those working in ITE in partnership with external stakeholders can, and have, sought to capture the complexity of ITE processes and outcomes and identify how to improve them. Next, the state of quality evaluation evidence collected by ITE programmes is in the U.S. is reviewed and then two examples of programme-specific models of quality assessment are discussed in detail. These programme-specific models differ in context and orientation and provide contrasting ideas on how others may develop their own context-specific frameworks.

4.2.2 Practitioner-led, Improvement-Oriented ITE Quality Assessments in the U.S.

How do practitioner-led efforts to evaluate quality among ITE programmes themselves differ from those instigated at the behest of policy-makers? And what can we learn from the more collaborative and context-specific approaches taken by other ITE programmes? To begin with, while VAM approaches to assessing teachers and ITE effectiveness are increasingly endorsed among American policy-makers, ITE programmes do not typically assess their graduates in this way. A U.S. survey of ITE across 240 institutions, conducted by the American Association of State Colleges and Universities, found that evidence of programme effectiveness was gathered through four primary methods: (a) observation systems supported, more or less, by faculty-developed

rubrics and programme standards from professional associations, (b) surveys of mentor teachers, school principals, and programme graduates, both during the programme and in follow-up data collection, (c) work samples and/or portfolios of teacher candidates, most often developed during methods courses or student teaching; and (d) state teacher certification tests (Wineburg, 2006). While few programmes reported using test scores to assess the effectiveness of their teacher candidates or graduates, all of the ITE programmes surveyed indicated that they use multiple sources of evidence to assess classroom performance. These sources of evidence included pre-testing and post-testing of students, teacher work samples and practicum-related assignments, surveys, interview data, classroom observations and videotape analysis among other things.

The survey also highlighted that most evidence collected by programmes was not longitudinal in nature. Only half reported that they tracked their graduates in the years following their preparation. Among this half, some programmes gathered data only once while others collected longitudinal data at year 1, 3 and 5. Notably, only one programme tracked graduates for more than 20 years. Still, nearly all respondents indicated that their programmes solicited judgments about their programmes from a wide variety of sources, including its graduates, placement school staff, colleges of arts and sciences, and other local and state advisory boards, accreditation organizations, and/or professional associations.

The survey of ITE programmes did not, however, ask how data were analysed by institutions or on what basis such evidence demonstrated effectiveness. The survey did enquire about how ITE programmes ensure the validity and reliability of their effectiveness measures, especially considering the wide variety of stakeholders concerned about ITE quality (e.g., schools, parents, policy-makers). While some responded that they did not assess for validity and reliability or were in the process of developing such checks, others described using a range of methods to validate their instruments. These methods ranged from simple methods such face validity, content validity, or construct validity to the more sophisticated such as inter-rater reliability, triangulation of data, or correlation studies. The survey data cannot, however, provide a glimpse into exemplary ways ITE faculty are assessing and reforming their own programmes. Thus, we now present a closer examination of the efforts at two universities – Boston College and Stanford – to evaluate and improve their own ITE

provision. This examination should provide valuable insights into how others might potentially go about such an endeavour.

The Boston College Model for ITE Assessment and Accountability

Starting in 2003, a team of inter-disciplinary teacher educator practitioners and researchers at Boston College began to systematically study ITE at the institutional level as part of a funded project called "Teachers for a New Era" (TNE) (see Ludlow *et al.*, 2008). Based on the work from this group, faculty of the Lynch School of Education at Boston College developed and implemented their own institution-specific and programme-level ITE assessment and accountability model (Ludlow *et al.*, 2010). This model, which took approximately six years to develop, was created in response to national calls for greater evidence of students' learning outcomes in American higher education institutions.

Before examining the model, it is important to highlight the institutional context in which this assessment model developed. Boston College (BC) is a medium-sized university with approximately 15,000 undergraduate and graduate students. Within it, the Lynch School of Education prepares approximately 250 to 270 undergraduate and graduate students to become teachers each year. The School has an explicit focus on teaching for social justice in its ITE mission, a commitment that no doubt reflects the city's historical struggle with racial integration and discrimination in its public schooling system. This commitment to social justice focuses the ITE programme on both the learning of teachers as well as students. Consequently, in addition to their methods courses and practicum, BC's teacher candidates take courses on the purposes and social contexts of schooling as well as courses in human development and how to address the various needs of a diverse student population (e.g., English-language learners, students with special needs, etc.). Relatedly, teacher candidates practice teaching in at least one school with a diverse student population and complete projects assessing the impact they had on pupils, and the implications for their teaching in the pursuit of social justice.

In this programmatic context, BC's education faculty developed an assessment and accountability model that had four key components (Ludlow *et al.*, 2010, p. 356 with emphasis added):

(1) *a conceptual framework* in which to locate a complementary portfolio of multiple studies that assess relevant processes and outcomes;

(2) *the involvement of faculty and relevant stakeholders* in order to change the culture of decision-making and interpretation;

(3) *measurements and assessments* that reflect the missions, goals, and values of the programme and the institution; and

(4) *the integration of the results* of various measures and assessments into a comprehensive data system linked to other databases.

In developing the first component – *a conceptual framework* rooted in a body of research that "assess relevant processes and outcomes" – Ludlow and his colleagues adopted the following framework, developed and described by members of the earlier TNE project:

> [The framework] represents the core aspects of teacher preparation and learning to teach that the ET concluded would have to be taken into account to understand teacher education's impact: (1) the characteristics of entering teacher candidates; (2) how these characteristics interact with the learning opportunities available in the programme; (3) how teacher candidates experience and make sense of these opportunities; (4) whether and how teacher candidates/graduates actually use what they learn in classrooms and schools (including teachers' strategies, interpretive frameworks, and ways of relating to students and others); (5) desired school outcomes, including pupils' academic, social, and civic learning as well as teacher retention and teaching for social justice; and, (6) how all of these are embedded within varying institutional, school, social, cultural, and accountability contexts and influenced by the differing conditions in which teachers work.
> (Cochran-Smith and Boston College Evidence Team, 2009, p. 460 as quoted
>
> by Ludlow et al., 2010, p.357, with numbers added)

Working from this framework, graphically represented by Figure 1 (Ludlow et al. 2010, p. 357), the BC faculty developed a number of studies to be conducted and selected a variety of measures through which to assess various parts of the framework. These studies and instruments represented a mix of quantitative, qualitative, and mixed-methods approaches to assessment. A wide variety of study designs and instruments were utilized due to the faculty's common conviction that no single measure could capture the complex processes and outcomes of ITE.
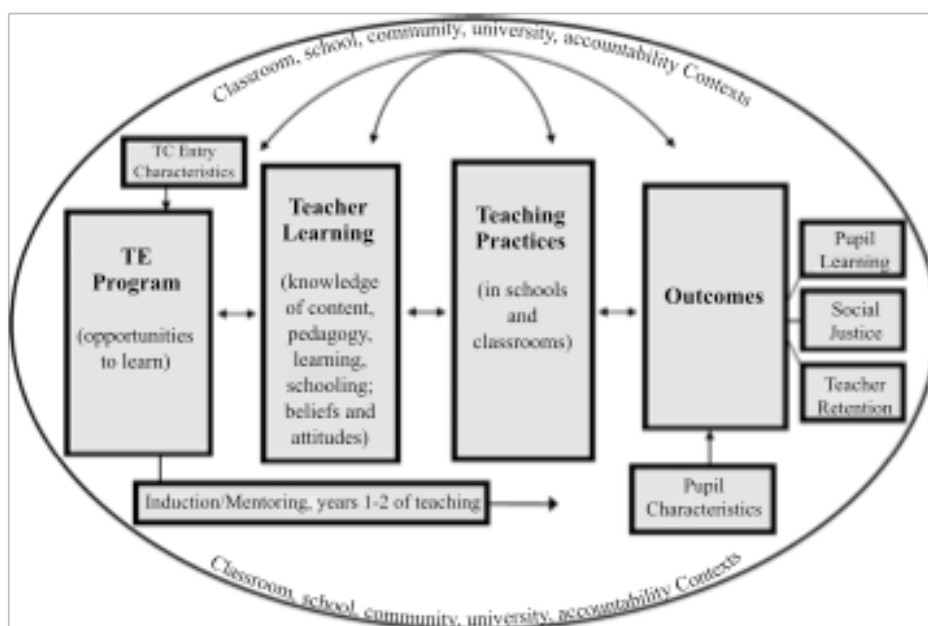
Figure 1. A conceptual framework for assessing teacher education. Created by Marilyn Cochran-Smith and the Boston College TNE Evidence Team in 2004 and used with permission.

The evidence portfolio has seven major projects, each of which aims to measure overlapping components of the framework and the relationship between them. The seven projects, or portfolio studies, are:

(1) a series of surveys examining teacher candidates'/graduates' perceptions, experiences, beliefs, and reported practices;

(2) a set of instruments that conceptualize and measure learning to teach for social justice as an outcome of teacher education;

(3) qualitative case studies, examining relationships among candidates' entry characteristics, learning in the programme, classroom practices, pupils' learning, and social justice;

(4) two analyses, drawing on longitudinal data bases from (1) and (3) above, designed to identify key interrelationships between teacher development and teacher retention;

(5) cross-sectional and value added assessment of the impact of BC graduates on pupils' test performance;

(6) comparison of graduates' classroom practices and pupils' performance on content tests for teachers from BC and from an alternate pathway into teaching in the same school district; and

(7) a mixed methods study of teacher candidates' ability to raise questions, document pupils' learning, and interpret and alter classroom practice using classroom-based inquiry.

(Ludlow et al., 2010, p.358)

Figure 2 represents BC's portfolio of studies and instruments for assessment based on their adopted framework (Ludlow et al., 2010, p. 358).
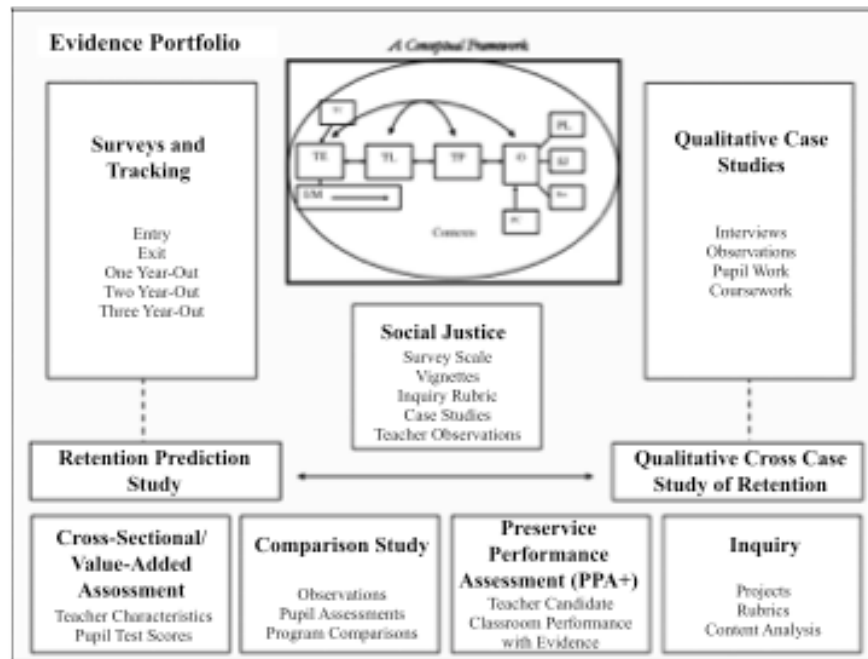

Figure 2.   Boston College TNE Evidence Portfolio. Cochran-Smith and the Boston College TNE Evidence Team (2009). Reproduced with kind permission from Sage © 2009.

The second component of the BC assessment model – the *involvement of faculty members and relevant stakeholders in culture change* – helps create a "data-rich environment that would support the development of a 'culture of evidence' in which the results of … the measures and assessments could inform decisions about initial teacher education" (Ludlow et al 2010, p. 359). To do this, the lead developers of the portfolio projects and studies shared drafts of the surveys and other data collection instruments with faculty from across ITE and gained their feedback to improve their work.

In addition, the lead investigators of the seven projects also organized "data workshops" with faculty from across the university and with colleagues fom partner schools. During these workshops, the researchers not only presented their collected data but, more importantly, actively engaged various stakeholders in interpreting and debating its significance. However, Ludlow and his colleagues found a number of

factors that impeded their progress in creating a "culture of evidence", namely finding a workable balance between those who view evidence as a "driver" of decision-making and those who see evidence as a basis for "informing" decision-making. The former perspective tends to promote a more positivist orientation that sees a direct and uncomplicated link between evidence and best practice or policy. Meanwhile the latter perspective stems from a more interpretive approach in which evidence is always seen as complex and situational and hence, evidence alone – without interpretation and an understanding of context – cannot drive decision-making.

The third component of the BC's model of ITE assessment and accountability – *the use of measurements and assessments that reflect the missions, goals, and values of the programme and the institution* – ensures that all the studies and instruments developed tied back to the Lynch School of Education's mission statement "to improve the human condition through education ... to expand the human imagination, and to make the world more just" (Ludlow et al., 2010, p. 360). More specifically, Ludlow and his colleagues made sure to link the BC model framework, its related studies, and measurement instruments back to one or more of ITE programme's five thematic goals: (1) *promoting social justice*, an overarching programmatic theme that encompasses the other four, (2) *constructing knowledge*, (3) *inquiring into practice*, (4) *meeting the needs of diverse learners*, and (5) *collaborating with others*. In practice, this meant translating the complex goal of "teaching for social justice" into a measurable outcome.

To do so, Ludlow and his team developed a set of "just measures", which are "tools, instruments, protocols, and studies" designed to document and assess the programme's progress, in various ways, toward this end (Cochran-Smith, Reagan and Shakman, 2009). The team conceptualized "learning to teach for social justice" as consisting of three "core" components:

1. a multi-faceted notion of students' learning as the goal of teaching,
2. a complex view of practice that includes teachers' knowledge, skills, interpretive stances, and learning in inquiry communities and
3. a political view of teaching as a player in maintaining or challenging inequities and incorporating or excluding the knowledge traditions of marginalized groups

(Cochran-Smith, Reagan and Shakman, 2009)

All three components are aimed toward the goal of "improving students' learning and enhancing their life chances" (Cochran-Smith, Reagan and Shakman, 2009). The team's

"just measures", for example, include a twelve-item, Likert-response "Learning to Teach for Social Justice-Beliefs scale" that was included in a suite of five surveys administered to both undergraduates and post-graduate students at specific points in time during the ITE programme and through their first three years after graduation (see Enterline *et al.*, 2008). This instrument "focuses exclusively on teacher candidates' and graduates' beliefs and perceptions" as related to teaching for social justice (Ludlow et al., 2010, p. 360). In this way, the BC ITE faculty emphasized that they strove to "measure what is valued" – the programme's own goals and values – rather than to simply "value what is measured", a common pitfall in research programmes that stems from limited time and resources and the need for simplification.

The fourth and final component of BC's model – *integration of multiple measures and assessments into a data system* – ensures that the large volume of quantitative and qualitative data generated each year from the model is systematically stored, integrated, and organized in a way that makes it easily accessible to various stakeholders at any time. To achieve this, BC's ITE faculty developed a "relational" data management system they named the Teacher Education System of Assessment (TESA). According to Ludlow et al. (2010), TESA contains interrelated tables that store a variety of information types, including master files of survey administrations, which link "the specific items used in each survey with the particular people who participated in each administration" (p. 361). In addition to collected data, TESA also contains and links data from other university databases (e.g., admissions, registration, practicum placement, and licensure results) with collected data on teacher candidates. Moreover, TESA features manuals, which are continually updated, on how to:

> (1) administer the surveys; (2) enter data into TESA; (3) perform routine statistical procedures; (4) generate routine reports for each survey administration; and (5) compile the survey results into a single continuous record extending over multiple years.
>
> (Ludlow et al., 2010, p. 362)

Since its launch, the TESA system has become institutionalized and its maintenance and management as well as its continual expansion and periodic reporting is now part of the role of a data manager and director of assessment on BC's faculty of education.

In sum, BC's model for ITE assessment and accountability is an example of a locally-developed and contextually-sensitive programme of quality measurement that

has been devised by ITE faculty in collaboration with other stakeholders across the university and in schools. The model's comprehensive conceptual framework and related set of assessment tools and studies maintain a continual focus on the programmatic theme of social justice, ensuring that each component individually and collectively provide evidence of the programme's overall effectiveness based upon its own goals and values. As a result, the BC model is arguably a quality assessment model that is strongly based on a particular value-infused mission. In contrast, the next quality framework examined below and developed by those at Stanford University represents a model that is more closely based around professional standards as reflected in its programme mission and measures.

The Stanford Teacher Education Program (STEP) Evaluation Framework

Another approach to assessing the quality of ITE and its ability to produce effective teachers is exemplified by work led by Linda Darling-Hammond at the Stanford Teacher Education Program (STEP) at Stanford University in California. As a teacher education expert and leading proponent of the professionalization of teaching in the U.S., Darling-Hammond (2006a) documents the research and assessment strategies used to evaluate the outcomes of STEP during its period of redesign from 2001 to 2006. STEP is a twelve-month post-graduate ITE programme for future secondary school teachers through which candidates earn both a master's degree and a California teaching credential. Although it is a rather small programme, graduating between 60 to 75 new teachers each year, its goal of developing innovative and knowledgeable leaders for schools and districts increases the impact and influence of its programme and graduates. STEP is known for having a number of strengths, notably (1) the involvement of senior faculty throughout the programme, (2) an emphasis on content pedagogy and learning to teach reflectively, and (3) a year-long clinical experience running in parallel with course work (Darling-Hammond, 2006a, p. 121). However, following a rather critical evaluation in 1999, the STEP programme was redesigned to address the following weaknesses:

> (1) a lack of common vision across the program, (2) uneven quality of clinical placements and supervisions, (3) a fragmented curriculum with

inconsistent faculty participation and inadequate attention to practical concerns such as classroom management, technology use, and literacy development, (4) limited use of effective pedagogical strategies and modelling in courses, (5) little articulation between courses and clinical work, and (6) little connection between theory and practice.

(Darling-Hammond, 2006a, p. 121 with numbers added)

To address these shortcomings, the STEP faculty redesigned the programme by first adopting a core conceptual framework, rooted in professional teaching standards, which provided a common vision for the programme.

In STEP's revised conceptual framework, teachers are viewed as "reflective practitioners" and "strategic decision makers who understand the processes of learning and development" and "can use a wide repertoire of teaching strategies to enable diverse learners to master challenging content" (Darling-Hammond, 2006a, pp. 121-122). Underpinning the conceptual framework is the progamme's orientation toward teaching for social justice, which aims to produce future teachers with the cultural and contextual knowledge, relevant skills, and moral commitment to help students from all backgrounds succeed in school. Thus, the programme was redesigned to emphasise learning and teaching in diverse contexts, including how to address the literacy needs across the curriculum as well as the needs of English-language learner. As a result, the programme was also redesigned to emphasise ways of fostering student inquiry and cooperative learning, to further STEP's vision for its teachers.

In addition to revising its conceptual framework, the STEP programme changed its school placement practices and chose to only place in professional development schools that had both the capacity and commitment to support its teacher candidates in professional communities of practice whose vision and values were aligned with STEP's. In short, STEP faculty adopted a core conceptual framework and social justice orientation that connects the components of the programme – both informing and shaping course design, programme assessment, and clinical work. Meanwhile, to better integrate theory and practice, a "clinical curriculum" was developed to more clearly specify, cultivate, and assess teaching practices and behaviours using rubrics reflective of professional standards. Other reforms – such as the integration of technology use across the curriculum – were also put in place. Finally, Darling-Hammond (2006a) noted that the curriculum and assessment in courses and clinical work were revised to

better align them with professional teaching standards commonly used for certification and accreditation.

To assess whether these changes improved the quality of the STEP programme, studies and assessment tools were developed and used to measure the STEP's outcomes in terms of graduates' "*preparedness*, *practice*, and *effectiveness* in supporting student learning" (Darling-Hammond, 2006a, p. 122-123 emphasis added). Drawing upon the work of Cochran-Smith (2001), the STEP faculty decided to measure ITE outcomes in terms of (1) *candidate learning* gained through particular courses and experiences during the programme, (2) the *teaching performance* of the candidates during the programme and as beginning teachers, and (3) the *outcomes* of this performance for classroom pupils. Overall, the STEP assessments focus most heavily on evidence in the first category – *candidate learning*.

STEP's three-pronged framework for assessment are also strongly rooted in the sets of national standards for teaching developed by the National Board for Professional Teaching Standards, the Interstate New Teacher Assessment and Support Consortium (INTASC), and the Performance Assessment for California Teachers (PACT). These widely used standards have been adopted by more than 30 U.S. states for teacher certification, licensure, and/or accreditation. Thus, the STEP framework reflects Darling-Hammond's and others' view that evaluations of teachers' performance and the impact of ITE on teaching must be based on widely agreed upon professional standards, or conceptions, identifying what skills, knowledge base, and dispositions contribute to effective, high-quality teaching (Darling-Hammond and Bransford, 2005; Darling-Hammond, 2006a). STEP's use of national professional standards in both revising the curriculum and assessment of coursework and the teaching practicum as well as the assessment for ITE effectiveness, helped make the programme a more transparent and coherent whole. Using PACT standards, which are specific to California but derived from those promoted nationally by INTASC, also make the collected data and results more comparable to assessment evidence collected from other Californian ITE programmes on their teacher candidates' knowledge and skills.

Using this three-pronged approach to measuring STEP's outcomes, a number of strategies were used to collect both quantitative and qualitative data. To track the first outcome related to "preparedness" – *candidates' learning*, the faculty collected survey data and conducted interviews with all its teacher candidates during and after the STEP

programme. In the interviews, faculty explicitly asked teacher candidates about specific aspects of the programme and their perceived effects, which helped STEP faculty better understand "relationships between programme design decisions and student experiences" (Darling-Hammond, 2006a, p. 126). Together, the survey and interview data provided a view of how candidates felt about the knowledge and skills gained in the programme, their experiences and struggles during and after it, and their preparedness and efficacy as novice teachers. Darling-Hammond (2006a) highlights how the data also enabled STEP faculty to evaluate a redesigned course and a particular strand of courses in addition to the effects of the programme as a whole. Darling-Hammond (2006a) also reported that the "most useful" data was "graduates' differential feelings of preparedness along different dimensions of teaching" which, when triangulated with other sources of data, "were directly useful in shaping ongoing [programme] reforms" (p. 125).

It is worth mentioning that, at one point, STEP faculty also interviewed 23 teacher candidates who had at least one year of teaching experience prior to entering STEP. While these already-experienced candidates were not separated or treated differently from other candidates during the programme, their interviews shed light on the particular goals and areas of concern for these types of individuals and how the STEP programme may best support their learning. Relatedly but in another context, the needs of teacher candidates who were more mature in age was also highlighted by a recent study in Ireland that found such candidates had different expectations of their ITE programme, tended to seek out alternative sources of support than faculty, and experienced their programme differently than younger candidates (Clarke, Lodge and Shevlin, 2012). To further support ITE evaluation, Darling-Hammond (2006a) also suggests ITE programmes "collect data on what candidates and graduates actually do in the classroom and what influences their decisions about practice" (pp. 125-126). In her view, this data may be linked to evidence about preparation and candidate learning – a point made by Hammerness (2006) and returned to below when discussing measuring the *teaching practice*.

To assess further *candidates' learning* during and after the programme, STEP faculty also collected objective data derived from pre-tests, post-tests, performance assessments, observations of practice, and samples of candidates' work. Evidence from these sources complement candidates' and graduates' perceptual data. Notably, STEP

faculty used pre-tests and post-tests designed to assess growth in candidates' knowledge related to learning, development, teaching, and assessment from the beginning to the end of the programme. The tests used – the Test of Teaching Knowledge (TKK) – were externally developed by members of the INTASC consortium in collaboration with the Educational Testing Services. The tests included 26 constructed response items across four sections and included a case-study, documentary analysis, and multiple-part questions about aspects of learning needs and theory, student behaviour, and modes of instruction among other things. After analysing the results the TKK pre- and post-tests, STEP faculty identified seven question items that candidates were able to answer correctly upon entering the programme by reading the question carefully and inferring the correct answer. However, for the most part, faculty saw significant growth in knowledge on most other items, suggesting the tests represented a valid measure of assessment pending further refinement.

As mentioned above, STEP faculty also derived data from student teacher work samples. More specifically, faculty used a "novice/expert" thinking framework (Berliner, 1986; Clarridge and Berliner, 1991) to code and score candidates' curriculum case study project and related assignments. By assessing student work, faculty members were able to trace and document the progression of candidates' thinking and application of theory through their successive curriculum drafts and related analytical essays – thinking that moved from "naïve generations" common to novice teachers' thinking to "sophisticated, theory-based explanations of the issues at play in their cases" that is more characteristic of expert teachers' thinking. Moreover, the analysis of student work was also combined with data from interviews with course instructors and a sample of students. As a result, Darling-Hammond (2006a) reported that they were able to identify "certain aspects of the course pedagogy [that] were important in helping student teachers think like a teacher" (p.128), which included:

(1) reading theoretical works in conjunction writing cases,
(2) sharing cases with peer readers,
(3) receiving specific theoretically grounded, concrete feedback from instructors, and
(4) revising the case several times in response to feedback about important elements of the context and teaching as well as potential theoretical explanations for what occurred.

(Darling-Hammond, 2006a, p. 128)

In this instance, STEP's experience illustrates the potential value in developing a way to assess systematically student teachers' work and makes a case for further triangulating such data with interviews to understand how elements of course design influence candidate learning and potentially lead to higher-order thinking and planning for effective teaching.

In addition to assessing candidates' work, STEP faculty also gathered data through longitudinal observations of candidates' clinical practice. To do so in a systematic way, faculty members and school mentors used a detailed rubric to evaluate candidates' teaching progress. The rubric was developed and used by Californian ITE programmes and was based upon the California Standards for the Teaching Profession, which contained six professional teaching standards and specific indicators for each standard. The observation rubric provided STEP faculty and school staff with a common template for assessment that drew evaluators' attention to a range of specific teaching practices that could be judged on a scale ranging from novice to expert. The rubric also provided guidance on how to judge candidates' performing by defining what counted as novice, proficient, or expert practice.

These observations of student teaching proved particularly valuable for the STEP programme as the practicum spanned an entire school year, giving ample opportunity for teacher educators to track the progress in candidates' performance as well as assess their performance at the end of the programme as an outcome. The longitudinal observations, conducted using a common standards-based rubric, also served to support quality improvement by (1) helping teacher candidates and evaluators focus their efforts and clarify goals, (2) building a consensus between university ITE faculty and school staff on the meaning of rubric categories and scoring, making their judgments more consistent with each other as time went on, (3) highlighting the need for intensive, collective discussions among stakeholders to develop shared-meaning of what assessments mean to increase their validity, and (4) prompting discussions as to how to independently confirm the progress in teaching practice indicated by the observations.

Finally, after candidates graduated, STEP faculty also collected data on *candidates' learning* from employers, including their perceptions of graduates' preparedness to teach and their direct observations of graduates' practice. Interestingly, STEP faculty noted a trend in the data suggesting that graduates are more critical of their own abilities than their employers are, as employers tended to rate graduates more

positively than the teachers themselves did – a finding echoed by a study of seven exemplary ITE programmes (Darling-Hammond, 2006c). STEP faculty did not, however, survey or interview their graduates in subsequent years to see how their views of their ITE experience and learning might have changed – an idea Darling-Hammond recommends since other research has suggested that, as teachers evolve in their career, their views of the utility of their ITE knowledge and experience changes.

So far, all the assessment instruments discussed have been related to measuring the "preparedness" of STEP candidates in terms of *candidates' learning* – the first defined outcome. The second outcome of STEP's assessment framework, which is related to "practice", is graduates' *teaching performance*, which refers to how graduates apply and enact what they have learned in real-life contexts. To measure this outcome, STEP faculty conducted two types of studies that examined candidates' teaching practice at two different points: (1) the end of the programme during the time they teach independently in their clinical setting and (2) during their first year teaching in schools after graduating STEP.

Critical to the assessment of *teaching performance* was linking observational data of practice to STEP's programme intentions. Doing so provided a way to identify and separate the effects of the ITE programme from other influences on graduates' teaching practice such as the school context or personal experiences. As a study by Hammerness (2006) illustrated, an ITE programme's intentions, or goals, may be gleaned from a close analysis of course syllabi and other programme documents together with interviews of faculty members. With a better understanding of the intentions of the programme under review, Hammerness then observed teacher candidates and recent graduates of the programme in their teaching and recorded whether there was "little", "some", or "strong" evidence of the five key programme goals she had identified in their teaching. Following Hammerness's example, STEP faculty sought to ensure rubrics for assessing *teaching performance* were closely aligned and reflected its own programme intentions, and thus provided greater clarity and coherence across the programme as a whole.

While observations of graduates' teaching practice linked to specific programme goals provide a measure of the effectiveness of ITE, STEP faculty also used a type of teaching portfolio assessments – known at the Performance Assessment for California Teachers (PACT) – to evaluate graduates' teaching performance. Developed in the

early 2000s by a consortium of twelve California university ITE programmes (including Stanford), the PACT provides a common and systematic way to evaluate teaching practice and is used to as the basis for credentialing recommendations (see Falk, 2013; Merino and Pecheone, 2013; Wilkerson, 2015).   The PACT consists of a "teaching event" portfolio in the subject area in which teachers teach along with "embedded signature assessments" used in each ITE programme such as the development of curriculum segments, pupil case studies, or analysis of learning.

The "teaching event" portfolio – which is the core of what makes the PACT unique – is a collection of several entries, or artefacts, focused on a particular instructional segment lasting approximately a week. The entries show evidence of the planning, instruction, assessment and reflection around a particular teaching event and include (1) descriptions of teaching contexts, (2) lesson plans, (3) one to two videotaped instructional sessions with reflective commentary provided by the candidate, (4) samples of pupils' work, and (5) candidate's written reflections about instruction and student learning during the unit (Darling-Hammond, 2006a, p. 130-131). Darling-Hammond highlights that PACT'S teaching event portfolio is a form of assessment that puts pupil learning – rather than teacher behaviours and knowledge – at the centre of focus. She describes how, in various ways, the portfolios focus on evidence of student learning of defined objectives and require candidates to assess how well they achieved their lesson objectives for all students and ways to modify instruction to better support diverse learners.

By combining teacher event portfolio evidence with programme-specific assessments, the PACT model allowed STEP to piece together multiple perspectives and gain a better understanding of the *teaching performance* of their candidates and recent graduates and how they compare to other universities' teacher graduates. The PACT, combined with longitudinal observations, also helped STEP faculty pinpoint the strengths and weaknesses within its own programme.

Finally, the third and last defined outcome for the STEP programme was *outcomes* of teaching practice for pupil learning. In other words, the third assessment area was essentially graduates' effectiveness in the classroom measured strictly in terms of student learning. This remains the most difficult but perhaps most important area to assess as numerous factors affect how students learn inside and outside school. As a result, the STEP faculty has not yet fully measured this outcome but is working towards

this goal through an ongoing study of the practices and effectiveness approximately 300 to 400 new elementary school teachers from STEP and other Californian university-based ITE programmes. When completed, these studies on the *outcomes* of teaching will complement and inform the measures developed for assessing *candidate's learning* and *teaching performance*.

4.2.3 Comparing the Practitioner-Led Quality Framework Models of the U.S.

In highlighting the ITE quality assessment models developed and used by BC and Stanford model, the differences in their approaches becomes clearer. While both models were aligned to their own programmatic vision, the differences in their visions, and hence their programme designs and emphases, led to somewhat contrasting quality assessment models. The vision of the STEP programme was more centred around professional teaching standards with social justice treated as an important but not all-defining theme. This stands in contrast to BC's model that adopted social justice as a driving purpose behind their programme and sought to cultivate such ideals and commitments as well as practices across all their ITE components and in their graduates. Thus, the difference in *a standards-based* versus a *value-based* mission influenced each programme's design and led to distinct conceptualizations of quality and assessment measurement in each context. For this reason, Stanford's quality assessment model can be considered *more* standards-based while BC's could be characterized as *more* values-based. This characterization reflects the fact that professional standards and a social justice orientation were present in both programmes and evaluation models, yet captures the point that Stanford put professional standards at the core of its work and BC took social justice as its core theme.

Both models, however, highlight how relational approaches to assessment can provide deeper understanding of ITE processes and effects, as well as the learning and development of teacher candidates that are not captured by ITE evaluation approaches developed at the state-level which address mainly policy-oriented goals. However, such frameworks and approaches do not have to ask wider audiences for the sake of gaining a more thorough and nuanced understanding of ITE quality. A recent Australian project aimed at assessing the quality of ITE programmes succeeds at bridging the gap between practitioner interests and those of external stakeholders, including policy-makers. This

Australian project is highlighted next and further illustrates the diversity in scope and designs of quality evaluation measures for ITE. In contrast to both the Stanford and BC model, the Australian project takes a more *theoretically-informed* approach to assessing the effectiveness of its ITE programmes.

4.3 Measuring the Impact of ITE in Australia: The SETE Project

In many ways, the Australian context of ITE parallels, and is influenced by, the American debates and policies toward ITE (Mayer, Zhu and Zeichner, 2013). In both countries, ITE programmes operate under decentralized governance systems where ITE is viewed as a "policy problem" in need of fixing. Consequently, in response to questions regarding the role and value of ITE in the Australian context, teacher educators across several universities collaborated to carry out a large-scale, longitudinal study to "investigate the effectiveness of teacher preparation for early career teachers employed in diverse settings across Australia" (Mayer *et al.*, 2015, p. 24). The four-year study, which began in 2010, was entitled the Studying the Effectiveness of Teacher Education (SETE). The collaborative project was funded by the Australian Research Council in partnership with teacher training institutes and colleges and the Education Departments of Victoria and Queensland.

Through SETE, researchers aimed to provide valuable insight for policy-makers and others on the "character, impact and outcomes [of] teacher education in Australia" (Mayer *et al.*, 2015, p. 24). To do this, the researchers adopted a particular *theoretically-informed*, "spatial" approach to the design of their evaluation based upon the works of Soja (1996) and Lefebre (1991). Mayer and her colleagues developed this approach from the central point of view that teacher education, its outcomes, and its quality could not be singularly measured or defined. Instead, they started from the premise that there exists "multiple ways of thinking about, enacting and performing teacher education [that] involve different, albeit related, spatial practices" (Rowan *et al.*, 2015, p. 281). From a spatial perspective, they saw teacher education not as "a singular construct, but, rather, a set of representations, practices, and experiences that are socio-spatial and relational in their nature" (p. 281).

As a result, they chose a "trialectical" way of examining teacher education that accounts for the interplay of three spaces: (1) the *imagined*, or "conceived", space of

representing teacher education through policy-making, planning, and mapping, (2) the *real*, or "perceived", space of teacher education as enactment of professional education, and (3) the *lived* space of teaching practice that contains the elements of the previous two spaces but exceeds their determination. More specifically, the "conceived" space is one in which definitions, measures, concerns and questions, and performance monitors of teachers (and by extension, teacher educators) are articulated by those within the wider policy environment. In contrast, the "perceived" space "refers to teacher education as it is understood within the domains of teacher education itself, particularly by those who design and deliver teacher preparation programmes" (Rowan *et al.*, 2015, p. 283). Thus, it is a space primarily inhabited by teacher educators who produce and manage professional knowledge (as opposed to policy knowledge) and make judgements about what future teacher should know and be able to do. For them, what quality or effectiveness looks like in teacher education may differ from those views held in the conceived space of policy. Meanwhile, the "lived" space refers to "teacher education as it is understood by pre-service and in-service teachers during and after they have completed their teacher preparation program" (Rowan *et al.*, 2015, p. 285). In the lived space, academic knowledge, skills, and beliefs acquired through ITE become part of the "practical consciousness" of beginning teachers, though they are also free to resist, subvert, or even re-imagine those ideas and practices offered by the other two spaces.

With attention paid to the meanings, concerns, and interests of each space, the SETE researchers developed three central research questions related to effectiveness in ITE which would require investigation in each space. The three central, but broad, questions of the SETE study were:

1. How well equipped are graduates to meet the requirements of the diverse settings in which they are employed?

2. What characteristics of teacher education programs are most effective in preparing teachers to work in a variety of school settings?

3. How does the teacher education program attended impact on graduate employment destination, pathways and retention within the profession?

To address these questions for each of the three spaces, the SETE researchers planned a multi-perspective study that combined both longitudinal in-depth qualitative research with large-scale quantitative research. The data collection, which took place between 2011 and 2014, proceed through four key, overlapping stages:

- An initial mapping of the terrain of teacher education activity (courses and pathways into teaching)

- Four rounds of surveys of teachers who graduated in 2010 and 2011 (Two in 2012; 2013; 2014)

- Surveys of the principals of graduates: 3 rounds (Two in 2012; 2013)

- Intensive case studies of new graduates involving 30 case study schools and approximately 180 teachers with visits in 2012, 2013 and 2014.

(Rowan *et al.*, 2015, p. 289)

In the first stage, SETE researchers completed an initial mapping of teacher education programmes, compiling data on the various existing pathways into teaching and their duration, structure, and components; approaches toward integrating theory and practice; and admission criteria. Then, with contextual data of all the programmes at hand, the study team then focused on collecting data in two ways: (1) surveys of graduate teachers and their principals, and (2) qualitative case studies. The surveys collected data on all three of the research questions and also provided necessary data on school contexts. Meanwhile, the in-depth case studies of a sample of teachers utilized the concept of *backward mapping*, which as Mayer and her colleagues (2015) explain, meant that they "did not start from the inquiry into teacher preparation courses to trace their effects on the practices of beginning teachers but rather [they] started from the exploration of the teachers' everyday lives in schools and then traced their practices back to their pre- service professional learning in various teacher education courses" (p. 292). Unlike many studies in which qualitative and quantitative data is compiled separately then combined to interpret and analyse results, the SETE researchers used a recursive strategy in which initial surveys informed the case study questions which then informed the next round of surveys and so on. In this way, their research design used both survey and interview data alongside contextual mapping to investigate ITE quality and effectiveness.

Overall, SETE researchers tracked two successive cohorts of ITE graduates in 2010 and 2011 (totalling nearly 5,000 beginning teachers) across Victoria and Queensland in their first years in teaching in schools and investigated the graduates' perceptions of the effectiveness of their ITE in preparing them to teach. At the same time, the study investigated the approximately 1,000 principals who had employed the

new ITE graduates to assess their perceptions of the new teachers' effectiveness. As a result of the study, SETE researchers uncovered prevailing trends among ITE graduates that illustrated the impact and effectiveness of policy and ITE programmes respectively. For example, the study identified nine key areas in which ITE graduates felt best prepared and revealed the effects of school characteristics, teacher characteristics, and programme type and contents on graduates' perceptions of preparedness. Moreover, case studies revealed professional experience as the core of learning teaching, highlighting the importance of the practicum and identified strengths of ITE programmes such as high quality university teaching staff and valuable practicum experiences among other things (Mayer *et al.*, 2015). Finally, the study also revealed insights into teachers' future employment prospects and environment as well as current trends in teacher mobility and retention.

The Australian study, in contrast to the Stanford and BC models of ITE assessment, represents a theory-based way to conceptualise and design collaborative research across an ITE sector. It also represents a way to address the concerns and impact of policy-makers on ITE and on graduates in the teaching field. As a result, it is a model that presents insights and ideas for future studies of ITE quality nationally.

4.4 An International Cross-Country Study of ITE Quality: The TEDS-M

While all the ITE quality assessment models reviewed thus far focus on ITE in a single national (or sub-national) context, there are wider comparative studies of ITE quality in the literature. One example is the Teacher Education and Development Study (TEDS-M), which is the only cross-national study to investigate the preparation of future primary and secondary mathematics teachers (Döhrmann, Kaiser and Blömeke, 2012; Tatto *et al.*, 2012; Ingvarson *et al.*, 2013). The study, carried out from 2006–2009 with the support of the International Association for the Evaluation of Educational Achievement (IEA), represents yet another way in which ITE quality has been assessed in the literature. It is an important example as all the models thus far have profiled particular studies and frameworks from single country contexts.

The key research questions driving this international, comparative study centred on the links between ITE policies, institutional practices, and the subject and pedagogy knowledge of future maths teachers at the end of their pre-service education. The

researchers wanted to find out, essentially, how national policies and institutional practices influence the outcomes of mathematics teacher education. A total of seventeen countries participated in the TEDS-M study (Botswana, Canada, Chile, Chinese Taipei, Georgia, Germany Malaysia, Norway, Oman, Philippines, Poland, Russian Federation, Singapore, Spain, Switzerland, Thailand, and the U.S). Notably, some participating ITE programmes had not previously had a system of data collection for regular and long-term evaluation in place before this study.

The researchers collected quantitative data from approximately 22,000 future teachers from 750 programs in about 500 teacher education institutions. Survey data was also collected from nearly 5,000 ITE staff, including mathematicians, mathematics educators, and general pedagogy educators. These participants were selected as representative samples of future teachers and their educators. The TEDS-M data included both questionnaires and knowledge assessments of graduates, designed and developed collaboratively by the international group of researchers, to determine ITE programme outcomes (Döhrmann, Kaiser and Blömeke, 2012). Using quantitative analysis, the researchers were able to compare the cross-country data to determine the relationships between policy, practices, and outcomes of ITE. Upon later reflection, Tatto et al. (2016) support this evaluation model but suggest that it could be strengthened if assessments of teacher knowledge "were to be linked with novice teachers' practices, and with how these practices in turn support learning" (p.12).

While providing a comparative examination of the various characteristics of teacher education systems in the participating countries, the findings of the study revealed that certain "teacher education systems" (meaning the regulatory policies, programmes, and context of ITE) tended to produce future teachers with the most knowledge of mathematics and subject pedagogy. These more effective ITE environments shared a greater focus on policies aimed at achieving the following:

1. enabling the teaching profession to compete for high-ability secondary school graduates;
2. balancing teacher demand and supply;
3. ensuring a rigorous system of assessment/accreditation of teacher education programs;
4. and setting high standards for entry to the profession (i.e., gaining registration licensing) after graduation.

(Ingvarson *et al.*, 2013)

Furthermore, the study indicated that the entry requirements of ITE programmes, opportunities to learn before and during the programme, graduation requirements, and strong systems of quality assurance were all associated with expected outcomes (Tatto et al., 2012). These findings reflect debates and conclusions occurring nationally and internationally about the most successful processes and policies for promoting teacher quality. Still, the TEDS-M study also highlights the wide variation in ITE curricula, programme design, and outcomes within, as well as across, the participating countries. Although only focused on the preparation of future mathematics teachers, the study shows the potential value for cross-country research in ITE while also highlighting many of the challenges such research presents.

5. Final Insights into ITE Quality Research, Measures, and Models

5.1 Synthesising Findings and Linking with Teacher Recruitment & Retention

The literature on ITE quality reviewed here represents a synthesis as well as a sampling of what is occurring in the field. There were other studies and models that could have been profiled but the selection reviewed here purposefully represents the diversity and design of contmeporary ways in which ITE quality is being conceptualised and measured. The review has aimed to highlight the historical and political context shaping debates about the quality and impact of ITE and the teachers it produces, whilst recognising that the responsibilities and challenges foisted upon teachers and, by extension, teacher educators are increasing. In light of the need to understand *how* to best prepare effective teachers, ITE programme faculty have begun to develop collaborative and context-sensitive ways to measure their own quality and effectiveness. The myriad of different instruments and tools for assessment may be appropriated and combined in different ways to develop such frameworks.

This review has profiled six examples of ways stakeholders and researchers inside and outside of ITE are doing this. First, Darling-Hammond, Hammerness, and colleagues offer a potential way to assess how ITE programmes compare to a particular model of "powerful", or exemplary, ITE by assessing programmatic vision and coherence as well as the quality of its practicum. Secondly, various states across the U.S. exemplify how VAM methodologies are combined with more commonly used instruments to compare and assess the effectiveness of ITE programmes and their graduates across local ITE sectors. Thirdly, Boston College's value-driven programme and corresponding model of assessment offered a more sophisticated and complex way of determining how ITE was developing and impacting the knowledge, skills, and dispositions of its graduates in the field. Fourthly, the Stanford STEP evaluation model – based heavily around professional standards and centrally employing portfolio-type teacher assessments – represents a contrasting way in which a single programme may conceptualize and measure its impact. Fifthly, the Australian SETE project provided insight into how an ITE quality assessment may benefit from a theoretically-driven approach to research design and address the concerns and interests of individuals in the three overlapping "spaces" through which ITE is conceptualised, shaped, and experienced. Finally, an international study of how ITE programmes in 17 countries are

developing future maths teachers illustrates how researchers are exploring the impact of national contexts, along with programmatic structures and practices, on the quality of ITE.

Across these contexts, discussion of recruitment strategies and retention rates are important as they both shape ITE quality and its operating contexts, as well as being shaped *by* ITE quality (Yost, 2006; Stotko, Ingram and Beaty-O'Ferrall, 2007; Goldhaber and Cowan, 2014; Kelly and Northrop, 2015). Research indicates that recruitment and retention trends are likely shaped by policy incentives (Vegas and Umansky, 2005; Henry, Bastian and Smith, 2012), wider social status of the teaching profession (Labaree, 2010; Lankford *et al.*, 2014), and employment prospects and working conditions (Boyd *et al.*, 2005, 2008; Donaldson and Johnson, 2010). Still, the type of candidates ITE programmes attract (Hutchings *et al.*, 2006; Flores and Niklasson, 2014; Straubhaar and Gottfried, 2014) as well as how effectively they are prepared through their programme for a teaching career (Gu and Day, 2007; McIntyre, 2010; Feiman-Nemser, Tamir and Hammerness, 2014) represent two key ways in which ITE affects recruitment and retention in the profession. When assessing ITE quality, recruitment to ITE and retention in the field of teaching need to be important factors in considering the long-term impact of ITE quality. If ITE programmes educate effective teachers but they, by and large, do not stay in the profession then quality ITE loses some of its importance. On the contrary, if ITE can prepare and help the profession to retain effective teachers, ITE programmes will not only benefit both in terms of more time and resources with which to prepare the next generation of teachers but increased influence on the future of the profession.

5.2 Notable Gaps within the Literature

There do exist notable gaps within the research on ITE quality measures and frameworks that deserve some attention. This literature review identified two such gaps – factors that were often neglected, or virtually absent, in ITE quality frameworks and studies. These gaps concern *ITE-school-community partnerships*, and the *role of teacher educators* in contributing to ITE quality. First, the strength and degree of collaboration within ITE-school partnerships were often not factored in as a measure of ITE quality although such a factor has implications on the availability and quality of the

practicum. While the "powerful" ITE indicators takes into account and measures programme coherence across the its coursework *and* fieldwork while also highlighting the importance of collaborative ITE-school partnerships, the frequency with which university- and school-based staff met or the extent of their collaborative activities were not considered. Other policy-oriented and improvement-oriented models also neglected to identify schools with which universities had a strong or weak relationships, and how that impacted on the messaging and overall effectiveness of the programme. This is a particularly relevant area for quality assessment in contexts such as Scotland which are preparing for, or in the process of developing, alternative ITE pathways.

Relatedly, there is a push in the literature regarding ITE-school partnerships for the incorporation of more context-specific modes of ITE, particularly those which engage teaching candidates in learning more about the school and communities in which they will likely work (Yogev and Michaeli, 2011; Hallman, 2012; Zeichner *et al.*, 2016; Richmond, 2017). A number of urban residency programmes, which typically partner ITE programmes to district leaders to address local needs, are designed to provide such in-depth learning from, and exposure to, communities and their history (Berry, Montgomery and Snyder, 2008; Hammerness and Matsko, 2013; Klein *et al.*, 2013). This is becoming more important as pupil demographics become more diverse, meaning students come from communities that are often culturally distinct from those of teacher candidates.

Zeichner (2010) argues that when ITE equally values community knowledge (alongside academic and experiential knowledge) and communities are engaged in helping prepare the teachers that will likely teach their children, "new synergies" may be created "through the interplay of knowledge from different sources" (p. 95). Incorporating communities into ITE-school partnerships or developing ITE-community partnerships to parallel existing ITE-school ones are possibilities that should be considered when measuring the quality and effectiveness of ITE programmes. At the very least, more attention and data could and perhaps should be collected to illuminate and identify what makes for an effective, collaborative ITE-school partnership.

Secondly, teacher educators were nearly completely ignored in the ITE quality literature, a point touched upon earlier in the review of the six attributes of quality ITE (Feuer *et al.*, 2013). In the wider ITE literature, there exists a growing body of research on the knowledge, skills, and dispositions of teacher educators and their journeys into,

and experiences within, ITE (Dinkelman, Margolis and Sikkenga, 2006; Lunenberg and Willemse, 2006; Swennen, Shagrir and Cooper, 2009; Margolin, 2011; Menter, 2011; Goodwin and Kosnik, 2013; Goodwin *et al.*, 2014; Leite, Dourado and Morgado, 2016). However, this review found no evidence that ITE quality assessments took into account such factors. Instead, only simple measures of teacher educators' qualifications and employment status were typically taken into account. This seems to be an ironic oversight given that, as high-quality teachers are the most critical factors in student achievement, high-quality teacher educators would (by extension) be a significant factor in creating effective teachers. As Goodwin et al. (2014) point out, "Commonsense reasoning says that quality teacher education relies on quality teacher educators" (p.284). Yet, as Snoek and colleagues (2010) found from examining the European debate on teacher education through six influential policy documents on the topic (published between 2005 and 2008), teacher educators and their professionalism were given limited attention beyond the general concern for increasing their effectiveness.

Despite this lack of attention, ITE quality measurements could certainly recognize the important role of teacher educators and potentially devise ways to measure their knowledge and skills and/or how well-supported (or not) they are by the ITE programme. Useful insight could be gleaned from Goodwin and colleagues' (2014) study. They used an online survey and qualitative interviews with 293 current teacher educators to discover their views on what constitutes the knowledge base and foundational elements of their practices, how prepared they felt for their roles, and how their knowledge and experience can inform others. Notably, Goodwin et al. (2014) found that practicing teacher educators often feel unprepared to assume their role but able to offer helpful insight into ways to conceptualise quality teacher educator preparation. Additionally, Swennen and colleagues' (2008) concept of *congruent teaching* could help identify, measure, and improve teacher educators' practice. Congruent teaching refers to the ability of teacher educators to link their own teaching to theory and model the kind of teaching they are promoting among teacher candidates. They suggest such congruent teaching may critically improve the quality of ITE by helping teacher candidates to bridge the all-too-common theory-practice divide and learn how to effectively draw upon theory in their own practice. In their small-scale (n=3), in-depth study, Swennen and colleagues provided participants with a workshop about congruent teaching and presented strategies for implementing such practices. They then

examined whether, with this support, the participants were able to carry out congruent teaching, which they were. The researchers concluded the acquisition of a language enabling them to discuss congruent teaching helps teacher educators to overcome associated challenges.

There are other possible ways to account for whether ITE programmes are effectively supporting teacher educators in their work (and hence, enhancing programme quality). These may include a check for the presence of (1) defined selection criteria for entry into the teacher educator profession, (2) formal education courses or a master's programme for new teacher educators, (3) induction programmes for teacher educators, (4) the opportunity and encouragement of teacher educators' participation in regional or international networks (ideas adapted from Snoek et al., 2010, pp. 38-39). Alternative or complementary measures of teacher educator quality (as a component of ITE quality) could be surveys and/interviews into their role in collaborative partnerships, perceptions of preparedness, and perceptions of support from the wider programme among other things.

6. Conclusion

While the quality or effectiveness of ITE programmes are a common concern and focus of policy-makers, researchers, and other stakeholders in most countries, there has not yet emerged a common "best" way to document, analyse, measure and judge the impact of ITE. Still, a number of insights emerge from this literature review. First, there are clearly many question to consider in the development stage of one's own ITE quality assessment framework. These questions are helpfully summarized by Feuer et al. (2013, pp. 6-7), who recommend considering:

(1) What is the primary purpose of the ITE evaluation system?

(2) Which aspects of ITE matter most?

(3) What sources of evidence will provide the most accurate and useful information about the aspects of teacher preparation that are of primary interest?

(4) How will the measures be analysed and combined to make a judgment about program quality?

(5) What are the intended and potentially unintended consequences of the evaluation system for ITE and education more broadly?

(6) How will transparency be achieved? What steps will be taken to help users understand how to interpret the results and use them appropriately?

(7) How will the evaluation system be monitored?

While the studies and measures profiled in this literature offer a variety of innovative ways in which one may conceptualise an ITE quality framework, it is worth considering early on whether one's framework would be best primarily guided by shared values (the Boston College Social Justice Model), professional standards (the Stanford STEP Model), or a particularly theoretical approach (the Australian SETE Spatial-based Research Design). Consideration may be given to the ways in which (1) Scottish ITE programmes engage in and manage their school partnerships (and potentially in the future, community-partnerships), (2) ways in which their teacher educators contribute toward the quality of ITE, and (3) ways in which ITE programmes support the professional practice and development of teacher educators.

Fortunately, the MQuITE study is in the unique position of being a meso-level, longitudinal study of ITE quality and impact. This is not common in the literature as most internal studies of ITE quality led by teacher educators/researchers tend to focus on one particular programme or institution. Meanwhile, other studies, usually conducted by external evaluators such researchers or governments assess a larger volume of ITE programmes on a macro-level. The MQuITE study presents an opportunity for a meso-level analysis in which the eight universities providing ITE in Scotland collaborate to assess the impact and quality of their programmes individually, comparatively, and collectively. Furthermore, the five-year cohort study allows for the collection of large amount of quality data on a particular cohort that will provide valuable and varied insights into the quality, impact and future direction of Scotland's ITE.

7. Using this literature review in the MQuITE study

This literature review is global in scope, but inevitably, because we were only able to review documents published in English, it is partial in its coverage. It is clear, too, that the majority of large-scale published work focusing on studies of ITE quality derives from the US, where the educational, political and cultural context is quite different to that of Scotland. Many of the publications reporting on such efforts are, understandably, focused on sharing the particular details of one approach; publications analysing such approaches from a critical perspective are less commonly available. Therefore, in drawing on this literature review to inform our own work, we are mindful of the need to interpret the studies reported here in relation to the contexts in which they are set. Indeed, the literature review is a rich source on which to draw, but should in no way be interpreted as a recipe or instruction manual for the creation of a quality framework. We are grateful to the lead author, Emilee Rauschenberger for providing us with such a rich and detailed set of ideas with which to work as we move forward in designing our own collaborative and context-appropriate framework for assessing the quality of our ITE programmes.

8. References

Arnott, M. and Ozga, J. (2010a) 'Education and nationalism: the discourse of education policy in Scotland', *Discourse: Studies in the Cultural Politics of Education*, 31(3), pp. 335–350. doi: 10.1080/01596301003786951.

Arnott, M. and Ozga, J. (2010b) 'Nationalism, governance and policymaking in Scotland: The Scottish National Party (SNP) in power', *Public Money & Management*, 30(2), pp. 91–96. doi: 10.1080/09540961003665503.

Arnott, M. and Ozga, J. (2012) 'Education policy and social justice', *Social Justice*.

Barber, M. and Mourshed, M. (2007) *How the world's best-performing schools systems come out on top*. New York: McKinsey & Company.

Barber, M., Mourshed, M. and Whelan, F. (2007) 'Improving education in the Gulf', *The McKinsey Quarterly*, 3947. Available at: http://twitmails3.s3.amazonaws.com/users/284092221/42/attachment/imed07.pdf (Accessed: 10 May 2017).

Berliner, D. C. (1986) 'In pursuit of the expert pedagogue', *Educational researcher*, 15(7), pp. 5–13.

Berry, B., Montgomery, D. and Snyder, J. (2008) *Urban Teacher Residency Models and Institutes of Higher Education: Implications for Teacher Preparation*. Center for Teaching Quality with the Aspen Institute and the National Council for the Accreditation of Teacher Education (NCATE). Available at: https://eric.ed.gov/?id=ED503644 (Accessed: 1 May 2017).

Biesta, G. (2009). Good education in an age of measurement: on the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability, 21*(1), 33-46.

Blömeke, S. *et al.* (2008) 'Effectiveness of teacher education', *ZDM*, 40(5), pp. 719–734. doi: 10.1007/s11858-008-0096-x.

Borko, H., Liston, D. and Whitcomb, J. A. (2007) 'Genres of empirical research in teacher education.(Editorial)', *Journal of Teacher Education*, 58(1), pp. 3–11. doi: 10.1177/0022487106296220.

Boyd, D. *et al.* (2005) 'Explaining the Short Careers of High-Achieving Teachers in Schools with Low-Performing Students', *The American Economic Review*, 95(2), pp. 166–171.

Boyd, D. *et al.* (2008) *Who leaves? Teacher attrition and student achievement*. National Bureau of Economic Research Cambridge, Mass., USA.

Boyd, D. J. *et al.* (2006) 'Complex by Design: Investigating Pathways Into Teaching in New York City Schools', *Journal of Teacher Education*, 57(2), pp. 155–166. doi: 10.1177/0022487105285943.

Brouwer, N. and Korthagen, F. (2005) 'Can Teacher Education Make a Difference?', *American Educational Research Journal*, 42(1), pp. 153–224. doi: 10.3102/00028312042001153.

Caena, F. and Margiotta, U. (2010) 'European Teacher Education: A Fractal Perspective Tackling Complexity', *European Educational Research Journal*, 9(3), pp. 317–331. doi: 10.2304/eerj.2010.9.3.317.

Cairney, P. and McGarvey, N. (2013) *Scottish Politics*. 2nd edn. Basingstoke: Palgrave Macmillan.

Canrinus, E. T. *et al.* (2017) 'Coherent teacher education programmes: taking a student perspective', *Journal of Curriculum Studies*, 49(3), pp. 313–333. doi: 10.1080/00220272.2015.1124145.

Chatterji, M. (2005) 'Evidence on "What Works": An Argument for Extended-Term Mixed-Method (ETMM) Evaluation Designs', *Educational Researcher*, 34(5), pp. 14–24. doi: 10.3102/0013189X034005014.

Chetty, N., Friedman, J. and Rockoff, J. E. (2014) 'Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates'. Available at: https://dash.harvard.edu/handle/1/30749073 (Accessed: 10 May 2017).

Clarke, M., Lodge, A. and Shevlin, M. (2012) 'Evaluating Initial Teacher Education Programmes: Perspectives from the Republic of Ireland', *Teaching and Teacher Education: An International Journal of Research and Studies*, 28(2), pp. 141–153. doi: 10.1016/j.tate.2011.08.004.

Clarridge, P. B. and Berliner, D. C. (1991) 'Perceptions of Student Behavior as a Function of Expertise', *The Journal of Classroom Interaction*, 26(1), pp. 1–8.

Clift, R. T. and Brady, P. (2005) 'Research on methods courses and field experiences', in Cochran-Smith, M. and Zeichner, K. M. (eds) *Studying teacher education: The report of the AERA panel on research and teacher education*. Mahwah, New Jersey: Lawrence Erlbaum Associates, pp. 309–424.

Cobb, P. *et al.* (2003) 'Design Experiments in Educational Research', *Educational Researcher*, 32(1), pp. 9–13. doi: 10.3102/0013189X032001009.

Cochran-Smith, M. (2004) 'Ask a different question, get a different answer: The research base for teacher education', *Journal of Teacher Education*, 55(2), pp. 111–115.

Cochran-Smith, M. (2005) 'The New Teacher Education: For Better or for Worse?', *Educational Researcher*, 34(7), pp. 3–17. doi: 10.3102/0013189X034007003.

Cochran-Smith, M. and Boston College Evidence Team (2009) '"Re-culturing" teacher education: Inquiry, evidence, and action', *Journal of Teacher Education*, 60(5), pp. 458–468.

Cochran-Smith, M., Reagan, E. M. and Shakman, K. (2009) 'Just Measures: Social Justice as a Teacher Education Outcome.', *Teacher Education and Practice*, 22(3), pp. 237–263.

Cochran-Smith, M. and Villegas, A. M. (2015) 'Framing Teacher Preparation Research: An Overview of the Field, Part 1', *Journal of Teacher Education*, 66(1), pp. 7–20. doi: 10.1177/0022487114549072.

Coggshall, J. G., Bivona, L. and Reschly, D. J. (2012) *Evaluating the Effectiveness of Teacher Preparation Programs for Support and Accountability. Research & Policy Brief*. Washington, D.C.: National Comprehensive Center for Teacher Quality. Available at: https://eric.ed.gov/?id=ED543773 (Accessed: 5 May 2017).

Conway, P. F. (2013) 'Cultural Flashpoint: The Politics of Teacher Education Reform in Ireland', *The Educational Forum*, 77(1), pp. 51–72. doi: 10.1080/00131725.2013.739021.

Dale, R. (1997) 'The State and Governance of Education: An analysis of the restructuring of the state-education relationship', in Halsey, A. H., Brown, P., and Wells, A. (eds) *Education, culture, economy, society*. Oxford: Oxford University Press, pp. 273–282.

Darling-Hammond (2000) 'How Teacher Education Matters', *Journal of Teacher Education*, 51(3), pp. 166–173.

Darling-Hammond, L. (2006a) 'Assessing Teacher Education: The Usefulness of Multiple Measures for Assessing Program Outcomes', *Journal of Teacher Education*, 57(2), pp. 120–138. doi: 10.1177/0022487105283796.

Darling-Hammond, L. (2006b) 'Constructing 21st-Century Teacher Education', *Journal of Teacher Education*, 57(3), pp. 300–314. doi: 10.1177/0022487105285962.

Darling-Hammond, L. (2006c) *Creating Powerful Teacher Education: Lessons from Excellent Teacher Education Programs*. Jossey Bass.

Darling-Hammond, L. and Bransford, J. (2005) *Preparing Teachers for a Changing World: What Teachers Should Learn and be Able to Do*. Jossey Bass.

Darling-Hammond, L. and Lieberman, A. (2013) *Teacher education around the world: Changing policies and practices*. London, UK: Routledge.

DeAngelis, K. J., Wall, A. F. and Che, J. (2013) 'The Impact of Preservice Preparation and Early Career Support on Novice Teachers' Career Intentions and Decisions', *Journal of Teacher Education*, 64(4), pp. 338–355. doi: 10.1177/0022487113488945.

Department for Education (2010) *The Importance of Teaching - The Schools White Paper 2010 CM7980*. Norwich: Stationary Office.

Dinkelman, T., Margolis, J. and Sikkenga, K. (2006) 'From teacher to teacher educator: Experiences, expectations, and expatriation', *Studying Teacher Education*, 2(1), pp. 5–23.

diSessa, A. A. and Cobb, P. (2004) 'Ontological Innovation and the Role of Theory in Design Experiments', *Journal of the Learning Sciences*, 13(1), pp. 77–103. doi: 10.1207/s15327809jls1301_4.

Döhrmann, M., Kaiser, G. and Blömeke, S. (2012) 'The conceptualisation of mathematics competencies in the international teacher education study TEDS-M', *ZDM*, 44(3), pp. 325–340.

Donaldson, G. (2011) *Teaching Scotland's Future: Report of a review of teacher education in Scotland*. Edinburgh: Scottish Government.

Donaldson, M. L. and Johnson, S. M. (2010) 'The Price of Misassignment: The Role of Teaching Assignments in Teach for America Teachers' Exit from Low-Income Schools and the Teaching Profession', *Educational Evaluation and Policy Analysis*, 32(2), pp. 299–323.

Drew, P. (2016) *Evaluating Teacher Education Programs through Performance-Based Assessments*. Hershey, Pennsylvania, USA: IGI Global.

EIS (2017) *Emergency Motion Highlights Fast-Track Teaching Opposition*, *The Educational Institute of Scotland (EIS)*. Available at: http://www.eis.org.uk/public.asp?id=3704.

Enterline, S. *et al.* (2008) 'Learning to Teach for Social Justice: Measuring Change in the Beliefs of Teacher Candidates', *The New Educator*, 4(4), pp. 267–290. doi: 10.1080/15476880802430361.

Falk, B. (2013) 'Lessons from the Performance Assessment of California Teachers (PACT)', *The New Educator*, 9(1), pp. 1–2. doi: 10.1080/1547688X.2013.751308.

Feiman-Nemser, S., Tamir, E. and Hammerness, K. (eds) (2014) *Inspiring Teaching: Preparing Teachers to Succeed in Mission-Driven Schools*. Cambridge, Massachusetts: Harvard Education Press.

Feuer, M. J. *et al.* (2013) *Evaluation of Teacher Preparation Programs: Purposes, Methods, and Policy Options*. Washington, D.C.: National Academy of Education. Available at: https://eric.ed.gov/?id=ED565694 (Accessed: 10 May 2017).

Flores, M. A. and Niklasson, L. (2014) 'Why do student teachers enrol for a teaching degree? A study of teacher recruitment in Portugal and Sweden', *Journal of Education for Teaching*, 40(4), pp. 328–343. doi: 10.1080/02607476.2014.929883.

Furlong, J. *et al.* (2000) *Teacher Education in Transition: Re-forming Professionalism?* Open University Press.

Furlong, J., Cochran-Smith, M. and Brennan, M. (eds) (2011) *Policy and Politics in Teacher Education: International Perspectives*. 1st edn. Routledge.

Gansle, K. A., Noell, G. H. and Burns, J. M. (2012) 'Do Student Achievement Outcomes Differ Across Teacher Preparation Programs? An Analysis of Teacher Education in Louisiana', *Journal of Teacher Education*, 63(5), pp. 304–317. doi: 10.1177/0022487112439894.

Gargani, J. and Strong, M. (2014) 'Can We Identify a Successful Teacher Better, Faster, and Cheaper? Evidence for Innovating Teacher Observation Systems', *Journal of Teacher Education*, 65(5), pp. 389–401. doi: 10.1177/0022487114542519.

Gideonse, H. (1993) 'The governance of teacher education and systemic reform', *Education Policy*, 7(4).

Goldhaber, D. and Cowan, J. (2014) 'Excavating the Teacher Pipeline: Teacher Preparation Programs and Teacher Attrition', *Journal of Teacher Education*, 65(5), pp. 449–462. doi: 10.1177/0022487114542516.

Goldhaber, D., Liddle, S. and Theobald, R. (2013) 'The gateway to the profession: Assessing teacher preparation programs based on student achievement', *Economics of Education Review*, 34, pp. 29–44. doi: 10.1016/j.econedurev.2013.01.011.

Goldstein, D. (2009) 'Teaching and the Mircale Ideology', *The American Prospect*, 15 July. Available at: http://prospect.org/article/teaching-and-miracle-ideology (Accessed: 10 May 2017).

Goodwin, A. L. *et al.* (2014) 'What Should Teacher Educators Know and Be Able to Do? Perspectives From Practicing Teacher Educators', *Journal of Teacher Education*, 65(4), pp. 284–302. doi: 10.1177/0022487114535266.

Goodwin, A. L. and Kosnik, C. (2013) 'Quality teacher educators = quality teachers? Conceptualizing essential domains of knowledge for those who teach teachers', *Teacher Development*, 17(3), pp. 334–346. doi: 10.1080/13664530.2013.813766.

Grossman, P. *et al.* (2008) 'Constructing Coherence: Structural Predictors of Perceptions of Coherence in NYC Teacher Education Programs', *Journal of Teacher Education*, 59(4), pp. 273–287. doi: 10.1177/0022487108322127.

Gu, Q. and Day, C. (2007) 'Teachers resilience: A necessary condition for effectiveness', *Teaching and Teacher Education*, 23(8), pp. 1302–1316. doi: 10.1016/j.tate.2006.06.006.

Guarino, C. *et al.* (2015) 'A Comparison of Student Growth Percentile and Value-Added Models of Teacher Performance', *Statistics and Public Policy*, 2(1), pp. 1–11. doi: 10.1080/2330443X.2015.1034820.

Hallman, H. L. (2012) 'Community-based field experiences in teacher education: possibilities for a pedagogical third space', *Teaching Education*, 23(3), pp. 241–263. doi: 10.1080/10476210.2011.641528.

Hammerness, K. (2006) 'From coherence in theory to coherence in practice', *Teachers College Record*, 108(7), p. 1241.

Hammerness, K. (2013) 'Examining Features of Teacher Education in Norway', *Scandinavian Journal of Educational Research*, 57(4), pp. 400–419. doi: 10.1080/00313831.2012.656285.

Hammerness, K. and Klette, K. (2015) 'Indicators of Quality in Teacher Education: Looking at Features of Teacher Education from an International Perspective', in LeTendre, G. and Wiseman, A. W. (eds) *Promoting and Sustaining a Quality Teacher Workforce*. Emerald Group Publishing Limited (International Perspectives on Education and Society, 27), pp. 239–277. doi: 10.1108/S1479-367920140000027013.

Hammerness, K. M. (2014) 'Visions of Good Teaching: Variation Coherence, and Opportunity to Learn', in Feiman-Nemser, S., Tamir, E., and Hammerness, K. (eds) *Inspiring Teaching: Preparing Teachers to Succeed in Mission-Driven Schools.* Cambridge, Massachusetts: Harvard Education Press, pp. 103–122. Available at: https://eric.ed.gov/?id=ED568812 (Accessed: 7 June 2017).

Hammerness, K. and Matsko, K. K. (2013) 'When Context Has Content A Case Study of New Teacher Induction in the University of Chicago's Urban Teacher Education Program', *Urban Education*, 48(4), pp. 557–584. doi: 10.1177/0042085912456848.

Hanushek, E. A. (2011) 'The economic value of higher teacher quality', *Economics of Education Review*, 30(3), pp. 466–479. doi: 10.1016/j.econedurev.2010.12.006.

Harris, D. N. and Herrington, C. D. (2006) 'Accountability, Standards, and the Growing Achievement Gap: Lessons from the Past Half‑Century', *American Journal of Education*, 112(2), pp. 209–238. doi: 10.1086/498995.

Henrick, E., Cobb, P. and Jackson, K. (2015) 'Educational Design Research to Support System-Wide Instructional Improvement', in Bikner-Ahsbahs, A., Knipping, C., and Presmeg, N. (eds) *Approaches to Qualitative Research in Mathematics Education*. Springer Netherlands (Advances in Mathematics Education), pp. 497–530. doi: 10.1007/978-94-017-9181-6_18.

Henry, G. T., Bastian, K. C. and Smith, A. A. (2012) 'Scholarships to Recruit the "Best and Brightest" Into Teaching: Who Is Recruited, Where Do They Teach, How Effective Are They, and How Long Do They Stay?', *Educational Researcher*, 41(3), pp. 83–92. doi: 10.3102/0013189X12437202.

Hostetler, K. (2005) 'What Is "Good" Education Research?', *Educational Researcher*, 34(6), pp. 16–21. doi: 10.3102/0013189X034006016.

Hulme, M. and Kennedy, A. (2016) 'Teacher education in Scotland: consensus politics and "the Scottish policy style"', in Beauchamp, G. et al. (eds) *Teacher Education in Times of Change: Responding to challenges across the UK and Ireland*. Bristol: Policy Press, pp. 91–108.

Hutchings, M. *et al.* (2006) 'An evaluation of innovative approaches to teacher training on the Teach First programme: Final report to the Training and Development Agency for Schools', *London: Institute for Policy Studies in Education*, 31.

Ingersoll, R., Merrill, L. and May, H. (2012) 'Retaining Teachers: How Preparation Matters', *Educational Leadership*, 69(8), pp. 30–34.

Ingvarson, L. *et al.* (2013) *An Analysis of Teacher Education Context, Structure, and Quality-Assurance Arrangements in TEDS-M Countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M)*. International Association for the Evaluation of Educational Achievement. Available at: https://eric.ed.gov/?id=ED545244 (Accessed: 10 May 2017).

Jacob, B. A. and Lefgren, L. (2008) 'Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education', *Journal of Labor Economics*, 26(1), pp. 101–136. doi: 10.1086/522974.

Jyrhämä, R. *et al.* (2008) 'The appreciation and realisation of research-based teacher education: Finnish students' experiences of teacher education', *European Journal of Teacher Education*, 31(1), pp. 1–16. doi: 10.1080/02619760701844993.

Kane, T. J., Rockoff, J. E. and Staiger, D. O. (2008) 'What does certification tell us about teacher effectiveness? Evidence from New York City', *Economics of Education review*, 27(6), pp. 615–631.

Kelly, S. and Northrop, L. (2015) 'Early Career Outcomes for the "Best and the Brightest": Selectivity, Satisfaction, and Attrition in the Beginning Teacher Longitudinal Survey', *American Educational Research Journal*, 52(4), pp. 624–656. doi: 10.3102/0002831215587352.

Kennedy, A. and Doherty, R. (2012) 'Professionalism and partnership: panaceas for teacher education in Scotland?', *Journal of Education Policy*, 27(6), pp. 835–848. doi: 10.1080/02680939.2012.682609.

Kennedy, M. M. (2007) 'Defining a Literature', *Educational Researcher*, 36(3), pp. 139–147. doi: 10.3102/0013189X07299197.

Klein, E. J. *et al.* (2013) 'Finding a third space in teacher education: creating an urban teacher residency', *Teaching Education*, 24(1), pp. 27–57. doi: 10.1080/10476210.2012.711305.

Koedel, C. *et al.* (2015) 'Teacher Preparation Programs and Teacher Quality: Are There Real Differences Across Programs?', *Education Finance and Policy*, 10(4), pp. 508–534. doi: 10.1162/EDFP_a_00172.

Kopp, W. (2011) *Chance to Make History: What Works and What Doesn't in Providing an Excellent Education for All*. New York: PublicAffairs.

Kumashiro, K. K. (2015) 'Review of proposed 2015 federal teacher preparation regulations', *Boulder, CO: National Education Policy Center. Retrieved January*, 10, p. 2015.

Labaree, D. (2010) 'Teach for America and Teacher Ed: Heads They Win, Tails We Lose', *Journal of teacher Education*, 61(1–2), p. 48.

Ladson-Billings, G. (2006) 'From the Achievement Gap to the Education Debt: Understanding Achievement in U.S. Schools', *Educational Researcher*, 35(7), pp. 3–12.

Lankford, H. *et al.* (2014) 'Who Enters Teaching? Encouraging Evidence That the Status of Teaching Is Improving', *Educational Researcher*, p. 0013189X14563600. doi: 10.3102/0013189X14563600.

Lasley, T. J., Siedentop, D. and Yinger, R. (2006) 'A Systemic Approach to Enhancing Teacher Quality: The Ohio Model', *Journal of Teacher Education*, 57(1), pp. 13–21. doi: 10.1177/0022487105284455.

Lefebre, H. (1991) *The Production of Space*. Oxford: Blackwell.

Leite, L., Dourado, L. and Morgado, S. (2016) 'Initial Science Teacher Education in Portugal: The Thoughts of Teacher Educators about the Effects of the Bologna Process', *Journal of Science Teacher Education*, 27(8), pp. 873–893. doi: 10.1007/s10972-016-9492-1.

Ludlow, L. *et al.* (2008) 'From students to teachers: using surveys to build a culture of evidence and inquiry', *European Journal of Teacher Education*, 31(4), pp. 319–337. doi: 10.1080/02619760802420842.

Ludlow, L. *et al.* (2010) 'An accountability model for initial teacher education', *Journal of Education for Teaching*, 36(4), pp. 353–368. doi: 10.1080/02607476.2010.513843.

Lunenberg, M. and Willemse, M. (2006) 'Research and professional development of teacher educators', *European Journal of Teacher Education*, 29(1), pp. 81–98. doi: 10.1080/02619760500478621.

Mangiante, E. (2011). Teachers matter: Measures of effectiveness in low-income minority schools. *Educational Assessment, Evaluation and Accountability, 23*(1), 41-63.

Margolin, I. (2011) 'Professional Development of Teacher Educators through a "Transitional Space": A Surprising Outcome of a Teacher Education Program', *Teacher Education Quarterly*, 38(3), pp. 7–25.

Mayer, D. *et al.* (2015) *Studying the Effectiveness of Teacher Education (SETE) Study Final Report*. Geelong, Victoria, Australia: Deakin University, p. 213.

Mayer, D., Pecheone, R. and Merino, N. (2013) 'Rethinking teacher education in Australia: The teacher quality initiative', in *Teacher Education around the World: Changing policies and practices*. London: Routledge, pp. 111–129.

Mayer, D., Zhu, X. and Zeichner, K. (2013) 'The Continuing "Problem"of Teacher Education: Policy Driven Reforms and the Role of Teacher Educators', in *Preparing Teachers for the 21st Century*. Berlin, Heidelberg: Springer, pp. 39–52. Available at: http://link.springer.com/chapter/10.1007/978-3-642-36970-4_3 (Accessed: 6 June 2017).

McIntyre, J. (2010) 'Why they sat still: the ideas and values of long‑serving teachers in challenging inner‑city schools in England', *Teachers and Teaching*, 16(5), pp. 595–614. doi: 10.1080/13540602.2010.507968.

Menter, I. (2011) 'Four "academic sub-tribes", but one territory? Teacher educators and teacher education in Scotland', *Journal of Education for Teaching*, 37(3), pp. 293–308. doi: 10.1080/02607476.2011.588018.

Menter, I. (2017) *The role and contribution of higher education in contemporary education*. SCDE: Edinburgh.

Merino, N. and Pecheone, R. (2013) 'The Performance Assessment for California Teachers: An Introduction', *The New Educator*, 9(1), pp. 3–11. doi: 10.1080/1547688X.2013.751310.

Mihaly, K. *et al.* (2012) *Where You Come from or Where You Go? Distinguishing between School Quality and the Effectiveness of Teacher Preparation Program Graduates*. SSRN Scholarly Paper ID 2020722. Rochester, NY: Social Science Research Network. Available at: https://papers.ssrn.com/abstract=2020722 (Accessed: 20 April 2017).

Mooney, G. and Scott, G. (2012) 'Devolution, social justice and social policy: the Scottish context', in *Social Justice and Social Policy in Scotland*, pp. 1–23.

Mourshed, M., Chijioke, C. and Barber, M. (2010) *How School Systems Get Better*. McKinsey & Company. Available at: http://www.mckinsey.com/clientservice/Social_Sector/our_practices/Education/Knowledge_Highlights/How%20School%20Systems%20Get%20Better.aspx (Accessed: 9 December 2010).

Muñoz, M. A. and Chang, F. C. (2007) 'The Elusive Relationship Between Teacher Characteristics and Student Academic Growth: A Longitudinal Multilevel Model for Change', *Journal of Personnel Evaluation in Education*, 20(3–4), pp. 147–164. doi: 10.1007/s11092-008-9054-y.

OECD (2005) *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*. illustrated edition. Paris: Organization for Economic Cooperation & Development Publishing.

Ramirez, F. O. *et al.* (2006) 'Student Achievement and National Economic Growth', *American Journal of Education*, 113(1), pp. 1–29.

Reardon, S. F. (2013) 'The widening income achievement gap', *Educational Leadership*, 70(8), pp. 10–16.

Richmond, G. (2017) 'The Power of Community Partnership in the Preparation of Teachers', *Journal of Teacher Education*, 68(1), pp. 6–8. doi: 10.1177/0022487116679959.

Rockoff, J. E. and Speroni, C. (2010) 'Subjective and objective evaluations of teacher effectiveness', *The American Economic Review*, 100(2), pp. 261–266.

Rogers, G. (2011) 'Learning‑to‑learn and learning‑to‑teach: the impact of disciplinary subject study on student‑teachers' professional identity', *Journal of Curriculum Studies*, 43(2), pp. 249–268. doi: 10.1080/00220272.2010.521262.

Roth, R. A. (ed.) (1999) *The Role of the University in the Preparation of Teachers*. London: Falmer Press Ltd.

Rowan, L. *et al.* (2015) 'Investigating the effectiveness of teacher education for early career teachers in diverse settings: the longitudinal research we have to have', *The Australian Educational Researcher*, 42(3), pp. 273–298. doi: 10.1007/s13384-014-0163-y.

Sahlberg, P. (2011) 'The fourth way of Finland', *Journal of educational change*, 12(2), pp. 173–185.

Sahlberg, P. (2015) *Finnish Lessons 2.0: What Can the World Learn from Educational Change in Finland?, Second Edition*. New York: Teachers College Press.

Sahlberg, P. (2016) 'The Global Educational Reform Movement and Its Impact on Schooling', in Mundy, K. et al. (eds) *Handbook of Global Education Policy*. London: John Wiley & Sons, pp. 128–138.

Sanderson, I. (2011) 'Evidence-based policy or policy-based evidence? Reflections on Scottish experience', *Evidence & Policy: A Journal of Research, Debate and Practice*, 7(1), pp. 59–76. doi: 10.1332/174426411X553007.

Sandoval, W. (2014) 'Conjecture Mapping: An Approach to Systematic Educational Design Research', *Journal of the Learning Sciences*, 23(1), pp. 18–36. doi: 10.1080/10508406.2013.778204.

Sandoval, W. A. and Bell, P. (2004) 'Design-Based Research Methods for Studying Learning in Context: Introduction', *Educational Psychologist*, 39(4), pp. 199–201. doi: 10.1207/s15326985ep3904_1.

Scott, G. and Wright, S. (2012) 'Devolution, social democratic visions and policy reality in Scotland', *Critical Social Policy*, 32(3), pp. 440–453. doi: 10.1177/0261018312444420.

Scottish Government (2016a) *Delivering Excellence and Equity in Scottish Education: A Delivery Plan for Scotland*. Edinburgh: Scottish Government.

Scottish Government (2016b) *National Improvement Framework for Scottish Education: Achieving excellence and equity*. Edinburgh: Scottish Government.

Shi, X. and Englert, P. A. J. (2008) 'Reform of teacher education in China', *Journal of Education for Teaching*, 34(4), pp. 347–359. doi: 10.1080/02607470802401537.

Sleeter, C. (2014) 'Toward Teacher Education Research That Informs Policy', *Educational Researcher*, 43(3), pp. 146–153. doi: 10.3102/0013189X14528752.

Snoek, M., Swennen, A. and van der Klink, M. (2010) 'The Teacher Educator: A Neglected Factor in the Contemporary Debate on Teacher Education', in Hudson, B., Zgaga, P., and Åstrand, B. (eds) *Advancing Quality Cultures for Teacher Education in Europe –Tensions and Opportunities*. Umeå, Sweden: Umeå School of Education, Umeå University, pp. 33–48. Available at: http://idiprints.knjiznica.idi.hr/196/1/2010%20Domovi%C4%87,%20Vizek%20Vidovi%C4%87.pdf#page=13 (Accessed: 7 June 2017).

Soja, E. (1996) *Thirdspace: Journeys to Los Angeles and Other Real-and-Imagined Places*. Oxford: Blackwell.

Stockard, J. and Wood, T. W. (2016) 'The Threshold and Inclusive Approaches to Determining "Best Available Evidence": An Empirical Analysis', *American Journal of Evaluation*, p. 1098214016662338. doi: 10.1177/1098214016662338.

Stotko, E. M., Ingram, R. and Beaty-O'Ferrall, M. E. (2007) 'Promising Strategies for Attracting and Retaining Successful Urban Teachers', *Urban Education*, 42(1), pp. 30–51. doi: 10.1177/0042085906293927.

Straubhaar, R. and Gottfried, M. (2014) 'Who Joins Teach For America and Why? Insights Into the "Typical" Recruit in an Urban School District', *Education and Urban Society*, p. 0013124514541463. doi: 10.1177/0013124514541463.

Strong, M., Gargani, J. and Hacifazlioğlu, Ö. (2011) 'Do We Know a Successful Teacher When We See One? Experiments in the Identification of Effective Teachers', *Journal of Teacher Education*, 62(4), pp. 367–382. doi: 10.1177/0022487110390221.

Swennen, A., Lunenberg, M. and Korthagen, F. (2008) 'Preach what you teach! Teacher educators and congruent teaching', *Teachers and Teaching*, 14(5–6), pp. 531–542. doi: 10.1080/13540600802571387.

Swennen, A., Shagrir, L. and Cooper, M. (2009) 'Becoming a teacher educator: Voices of beginning teacher educators', in *Becoming a teacher educator*. Springer, pp. 91–102. Available at: http://link.springer.com/content/pdf/10.1007/978-1-4020-8874-2_7.pdf (Accessed: 7 June 2017).

Tatto, M. T. *et al.* (2012) *The Mathematics Teacher Education and Development Study (TEDS-M). Policy, practice, and readiness to teach primary and secondary mathematics: First findings*. Amsterdam: International Association for the Evaluation of Educational Achievement.

Tatto, M. T. *et al.* (2016) 'The Emergence of High-Stakes Accountability Policies in Teacher Preparation: An Examination of the U.S. Department of Education's Proposed Regulations', *education policy analysis archives*, 24(0), p. 21.

Valcke, M. (2013) '"Evidence-Based Teaching, Evidence-Based Teacher Education"(Quality of Teachers and Quality of Teacher Education)', in Zhu, X. and Zeichner, K. (eds) *Preparing Teachers for the 21st Century*. Berlin, Heidelberg: Springer, pp. 53–66. Available at: http://link.springer.com/chapter/10.1007/978-3-642-36970-4_4 (Accessed: 6 June 2017).

Vegas, E. and Umansky, I. (2005) 'Improving Teaching and Learning through Effective Incentives: Lessons from Education Reforms in Latin America', in *Incentives to Improve Teaching: Lessons from Latin America*. World Bank, pp. 1–19.

Walsh, E. and Isenberg, E. (2015) 'How Does Value Added Compare to Student Growth Percentiles?', *Statistics and Public Policy*, 2(1), pp. 1–13. doi: 10.1080/2330443X.2015.1034390.

Watson, C. (2010) 'Educational policy in Scotland: inclusion and the control society', *Discourse: Studies in the Cultural Politics of Education*, 31(1), pp. 93–104. doi: 10.1080/01596300903465443.

Whitty, G. (2006) 'Education(al) research and education policy making: is conflict inevitable?†', *British Educational Research Journal*, 32(2), pp. 159–176. doi: 10.1080/01411920600568919.

Wilkerson, J. R. (2015) 'The Need for Instructional Sensitivity and Construct Clarity in PACT: A Commentary on "Examining the Internal Structure Evidence for the Performance Assessment for California Teachers"', *Journal of Teacher Education*, 66(2), pp. 184–192. doi: 10.1177/0022487114561029.

Wineburg, M. S. (2006) 'Evidence in Teacher Preparation: Establishing a Framework for Accountability', *Journal of Teacher Education*, 57(1), pp. 51–64. doi: 10.1177/0022487105284475.

Wiseman, A. W. (2007) 'National school systems', in Bank, B. J. (ed.) *Gender and Education: An encyclopedia*. Westport, CT, US: Greenwood Press, pp. 201–208.

Xu, Z., Hannaway, J. and Taylor, C. (2007) *Making a Difference?: The Effect of Teach for America in High School*. Working Paper 17. The Urban Institute and National Center for the Analysis of Longitudinal Data in Education Research (CALDER). Available at: http://www.urban.org/publications/411642.html (Accessed: 4 January 2010).

Yogev, E. and Michaeli, N. (2011) 'Teachers as society-involved "organic intellectuals": Training teachers in a political context', *Journal of Teacher Education*, 62(3), pp. 312–324.

Yost, D. S. (2006) 'Reflection and Self-Efficacy: Enhancing the Retention of Qualified Teachers from a Teacher Education Perspective', *Teacher Education Quarterly*, 33(4), pp. 59–76.

Young, J., Hall, C. and Clarke, T. (2007) 'Challenges to university autonomy in initial teacher education programmes: The cases of England, Manitoba and British Columbia', *Teaching and Teacher Education*, 23(1), pp. 81–93. doi: 10.1016/j.tate.2006.04.008.

Zeichner, K. (1999) 'The New Scholarship in Teacher Education', *Educational Researcher*, 28(9), pp. 4–15.

Zeichner, K. (2010) 'Rethinking the Connections Between Campus Courses and Field Experiences in College- and University-Based Teacher Education', *Journal of Teacher Education*, 61(1–2), pp. 89–99. doi: 10.1177/0022487109347671.

Zeichner, K. *et al.* (2016) 'Engaging and Working in Solidarity With Local Communities in Preparing the Teachers of Their Children', *Journal of Teacher Education*, 67(4), pp. 277–290. doi: 10.1177/0022487116660623.

Zeichner, K. M. and Cochran-Smith, M. (2005) *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education*. 1st edn. Routledge.