

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries



by Maura Best
Pat Knight
Petra Lietz
Craig Lockwood
Dita Nugroho
Mollie Tobin

August 2013

This research was funded by the Australian Agency for International Development (AusAID). The research was commissioned as part of a joint call for systematic reviews with the Department for International Development (DFID) and the International Initiative for Impact Evaluation (3ie)

The views expressed are those of the authors and not necessarily those of the Commonwealth of Australia. The Commonwealth of Australia accepts no responsibility for any loss, damage or injury resulting from reliance on any of the information or views contained in this publication.

The authors are part of ACER and were supported by the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre).

The EPPI-Centre reference number for this report is 2104.

This report should be cited as:

Best M, Knight P, Lietz P, Lockwood C, Nugroho D, Tobin M (2013) The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries. Final report. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

ISBN: 978-1-907345-57-9

© Copyright

Authors of the systematic reviews on the EPPI-Centre website (<http://eppi.ioe.ac.uk/>) hold the copyright for the text of their reviews. The EPPI-Centre owns the copyright for all material on the website it has developed, including the contents of the databases, manuals, and keywording and data extraction systems. The centre and authors give permission for users of the site to display and print the contents of the site for their own non-commercial use, providing that the materials are not modified, copyright and other proprietary notices contained in the materials are retained, and the source of the material is cited clearly following the citation details provided. Otherwise users are not permitted to duplicate, reproduce, re-publish, distribute, or store material from this website without express written permission.

Contents

List of abbreviations	iv
1. Background	1
1.1 Aims and rationale for current review	1
1.2 Definitional and conceptual issues	1
1.3 Theory, policy and practice background	5
1.4 Research background	11
1.5 Authors, funders, and other users of the review	12
1.6 Review questions	13
1.7 Systematic review background	14
2. Methods used in the review	21
2.1 User involvement	21
2.2 Identifying and describing studies	22
2.3 In-depth review	26
3. Identifying and describing literature	30
3.1 Identification, screening and keywording literature	30
4. In-depth review: results	36
4.1 Characteristics of assessment programmes used in education policy-making	36
4.2 Goals and uses of assessment programmes	37
4.3 Stages of the policy process	38
4.4 Education policies resulting from the use of assessment data	39
4.5 Facilitators and barriers to the use of assessment data	41
4.6 Synthesis of the evidence	43
4.7 Types of policy-synthesis	49
5. Summary of results	62
6. Implications of the review	65
6.1 Implications for intended audience of the review	65
6.2 Strengths and limitations	65
6.3 Opportunities for further research	65
7. References	67
Appendices	74
Appendix 1.1: Authorship of this report	74
Appendix 1.2: Classification of developing countries	76
Appendix 2.1: Exclusion criteria in screening	78

Appendix 2.2: Database keywords and descriptors	79
Appendix 2.3: Data extraction tool with keywords	87
Appendix 2.4: Synthesis table	90
Appendix 3.1: Description of ‘academic papers with possible policy suggestions’ (n=78)	91
Appendix 3.2: Synthesis table of included records for detailed analysis (n=50)	95
Appendix 4.1: Characteristics of assessment programmes in education policy- making	100
Appendix 4.2: Goals and uses of assessment programmes	105
Appendix 4.3: Stages of the policy process	106
Appendix 4.4: Education policies	107
Appendix 4.5: Facilitators and barriers	109

List of abbreviations

3ie	International Initiative for Impact Evaluation
ACER	Australian Council <i>for</i> Educational Research
AEI	Australian Education Index
AJOL	African Journals Online
AusAID	Australian Agency for International Development
BEI	British Education Index
BLDS	British Library for Development Studies
CONFEMEN	La Conférence des ministres de l'éducation des pays ayant le français en partage
CREATE	Consortium for Research on Educational Access, Transitions and Equity
DOAJ	Directory of Open Access Journals
DFID	Department for International Development (UK)
EPPI-Centre	Evidence for Policy and Practice Information and Co-ordinating Centre (UK)
ERC	Education Research Complete
ERIC	Education Resources Information Center
ETS	Educational Testing Service (USA)
IAEP	International Assessment of Educational Progress
ICT	Information and computer technology
IEA	International Association for the Evaluation of Educational Achievement
IEA CIVED	IEA Civic Education Study
IEA ICCS	IEA International Civic and Citizenship Study
IMF	International Monetary Fund
JBI	Joanna Briggs Institute (AUS)
LAMJOL	Latin American Journals Online
LLECE	Latin American Laboratory for Assessment of the Quality of Education
MLA	Monitoring Learning Achievement (UNICEF)
NRC	National Research Coordinator (World Bank)
OECD	Organisation for Economic Co-operation and Development
PASEC	Le Programme d'Analyse des Systèmes éducatifs des États et gouvernements membres de la CONFEMEN (Programme on the Analysis of Education Systems)

PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
PREAL	Partnership for Educational Revitalization in the Americas
PREL	Pacific Resources for Education and Learning
RTI	Research Triangle Institute (USA)
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
SCIELO	Scientific Electronic Library Online
SERCE	Second Regional Comparative and Explanatory Study
SES	Socio-economic status
STEM	Science, technology, engineering and mathematics
TALIS	Teaching and Learning International Survey
TERCE	Third Regional Comparative and Explanatory Study
TIMSS	Trends in International Mathematics and Science Study
TIMSS-R	Trends in International Mathematics and Science Study - Repeat
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNESCO-IIEP	United Nations Educational, Scientific and Cultural Organization - International Institute for Educational Planning
UNICEF	United Nations Children's Fund

1. Background

Outline of chapter

This chapter provides the background to the systematic review. It aims to set the review within a context of theory, policy and practice, and reviews prior research relevant to the topic. Section 1.1 introduces the basic principles that are discussed in more detail in the rest of the chapter. This is followed by a set of definitions and concepts that are central to the review topic in Section 1.2.

Section 1.3 then grounds the review in existing theories, policies and practices that are pertinent to the topic: the dramatic rise of assessment programmes within developing countries, the concept of evidence-based policy-making, and the different uses of assessment to serve as evidence. The next section reviews existing research, including relevant systematic reviews, on the topic.

The final sections present the main and potential users of the review and outline the questions that are addressed by this review.

1.1 Aims and rationale for current review

There is a documented global rise in the number of countries undertaking national learning assessments (Benavot and Tanner 2007, UNESCO 2008), as well as international and regional learning assessments (Kamens and McNeely 2009). Much of this growth, especially in national learning assessments, has occurred in economically developing countries.

Little is known, however, on how these assessments affect education policy and practice in developing countries.

This review examined the impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries. This particular focus on policies regarding resources and teaching and learning practices stemmed from an observation that, particularly in economically developing countries, analyses of data from such assessments are used to make policy recommendations in those areas (e.g. Abulibdeh and Abdelsamad 2008, Assessment and Evaluation Center 2006).

This review synthesised evidence by employing a framework synthesis approach to accommodate the anticipated diverse types and quality of literature. The use of an initial conceptual framework effectively guided analysis to consider established evidence as well as policy considerations. At the same time, the use of a preliminary conceptual framework allowed for the development of new evidence to emerge, as on a global scale, little is known about the impact of these assessment programmes in developing countries.

Therefore, the results of this review will inform relevant stakeholders who are involved in the planning, funding and in using data from these assessments as to the types of policy impacts found in developing countries. Furthermore, the results of this review will help to guide use of assessment data and participation in assessment programmes.

1.2 Definitional and conceptual issues

For the purposes of this review, the concepts embedded within the review title, *The impact of national and international assessment programmes on education*

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries, are elaborated and defined below.

1.2.1 National, regional and international assessment programmes

International assessments were initially conceived to explore cross-national variation in educational institutions and processes and their relationship to student learning outcomes (Keeves 1995), and many countries have used national assessments for entry to further education (Greaney and Kellaghan 2008, Kamens and McNeely 2009). Currently, national and international assessment programmes are mainly used to monitor and evaluate the quality of student learning outcomes (Postlethwaite and Kellaghan 2008) and are designed to enable comparisons over time (Greaney and Kellaghan 2008), although in some countries national assessments still serve a 'gate-keeping' function where there are limitations to the availability of places in subsequent stages of schooling.

A common understanding appears to be that the main aim of conducting standardised learning assessments is to provide information on a country's educational outcomes, which, in turn, assists policy-makers and other stakeholders in the education system with making policy and resourcing decisions for improvement (Benavot and Tanner, 2007, Braun and Kanjee 2006, Forster 2001), although the appropriateness of using standardised tests in this way has been questioned (e.g. Goldstein and Thomas 2008, Popham 1999).

In order to provide information regarding educational outcomes, national and international assessments are designed to be standardised cognitive assessments, which provide evidence about the level of student achievement in identified curriculum areas, according to Postlethwaite and Kellaghan (2008). The term 'standardised', in this context, usually refers to consistency in test design, content, administration and scoring to ensure comparability of the results across students and schools (deLandshere 1997). The curricular areas mainly assessed in international assessments - mathematics, language, science and civic and citizenship education - are principally the same curricular areas assessed in national assessments, as these subjects constitute the majority of curricula in primary education cross-nationally (Kamens and McNeely 2009).

Furthermore, a third type of assessment programme, regional assessments, have been undertaken to compare samples of schools in a region of the world in which countries may share similar economic and social conditions (Kamens and Benavot 2011) in order to explicitly compare student achievement cross-nationally. In this review, regional assessments were understood to be a distinct type of international assessment programme.

For this review, national, regional and international assessment programmes were understood to be assessment programmes conducted in primary and secondary education, and included assessment programmes that were undertaken at the sub-national level (e.g. state level). Sub-national assessments were understood to be standardised large-scale assessments, often implemented in countries with decentralised education systems (e.g. India). References to non-standardised assessments were not considered in this review.

1.2.2 Education policy

The review sought to understand the impact of national and international assessment programmes on education policy within developing countries, as there has been an increased focus for educational planning to improve understanding of education policy-making processes (Haddad 1995).

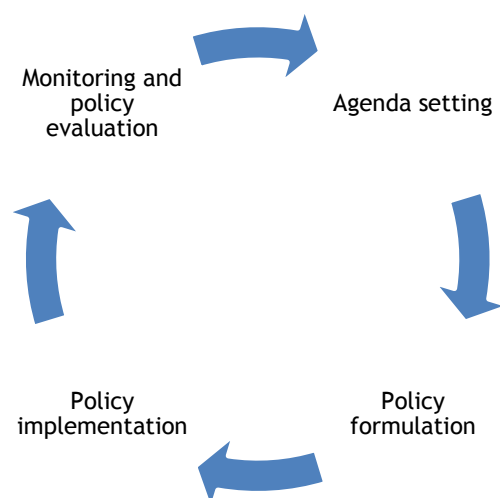
In a United Nations Educational, Scientific and Cultural Organization-International Institute for Educational Planning (UNESCO-IIEP) booklet from the series 'Fundamentals in Educational Planning', Haddad (1995) provides a useful definition of policy-making:

An explicit or implicit single decision or group of decisions which may set out directives for guiding future decisions, initiate or retard action, or guide implementation of previous decisions. (Haddad 1995, p. 18)

Educational policies may be concerned with content, instruction, resources and assessment. At the system level, these policies target educational issues such as curriculum development, the allocation of resources in education, the use of learning assessments and the development of achievement standards, as well as standards of teacher qualifications and teaching and learning practices, among others.

A framework that is commonly employed in discussions of the policy-making process is the concept of the policy cycle, in which the process is seen as having separate stages. A number of models of the policy cycle have been proposed, generally involving six to eight stages (Bridgman and Davis 2004, Haddad 1995, Young and Quinn 2002). This review discussed education policy-making by using a simplified policy cycle model from Sutcliffe and Court (2005) (Figure 1.1).

Figure 1.1: Simplified model of the policy cycle; source: Sutcliffe and Court (2005)



In more detail, these four stages are:

- Agenda setting: awareness of and priority given to an issue or problem;
- Policy formulation: the ways (analytical and political) options and strategies are constructed;
- Policy implementation: the forms and nature of policy administration and activities on the ground; and
- Monitoring and policy evaluation: the nature of monitoring and evaluation of policy need, design, implementation and impact.

Although learning assessments themselves form part of the monitoring and policy evaluation stage, data from assessments can be used at different stages of the

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

process, and a later section of this document describes the different issues to be taken into consideration for the use of data at each stage.

1.2.3 Developing country

This review used the Australian Agency for International Development's (AusAID's) list of developing countries as declared by the Minister for Foreign Affairs for the purposes of the Overseas Aid Gift Deduction Scheme established by the Income Tax Assessment Act 1997 (last updated in July 2009). The list includes 150 countries within eight world regions (Appendix 1.2).

1.2.4 Resource allocation

Resource allocation refers to the resources that schools receive, which are frequently defined as inputs (Hanushek 2003). Inputs may include traditional measures of resources such as expenditure per student (Hanushek and Kimko 2000) or national education budgets. Resources may include instructional materials (e.g. textbooks), school supplies (e.g. pencils), equipment (e.g. audio-visual equipment) and facilities (e.g. heating and cooling systems) (Mullis et al. 2005), and also class- and school-level characteristics such as class size (Krueger 2002), teacher-to-student ratios and instructional time (Woessmann 2000). Resources may refer to teacher characteristics such as teacher experience and level of teacher qualifications (Woessmann 2000). Furthermore, this concept refers not only to resources that are at the discretion and within the decision-making powers of the school. Instead, it includes all resources, monetary, human and physical, that are included in a country's education budget and for which the allocation or decision-making powers may rest with various levels of a country's administration (OECD 2010).

1.2.5 Teaching and learning practices

In order to improve student learning outcomes, there is a focus on improving school- and classroom-level factors such as teaching and learning strategies. Teaching and learning practices are more amenable to being affected by policy interventions than other factors related with student learning outcomes, such as student, family and community background characteristics (Hattie 2009, OECD 2009a). Holistically, teaching has been conceptualised by Hattie (2009) as a process which requires content knowledge, the ability to guide learning through planned interventions for specified outcomes using a wide variety of strategies, the monitoring of student and self-learning, and the ability to provide a range of feedback on student learning.

Teaching practices have been internationally operationalised in the Organisation for Economic Co-operation and Development's (OECD's) Teaching and Learning International Survey (TALIS). Using a policy framework, 24 participating countries jointly developed indicators of teaching practices (OECD 2009a). This thematic framing of teaching practices includes classroom management and discipline, practices with a student orientation such as differentiated learning and student support, and enhanced learning activities which require higher-order thinking. Teaching practices may furthermore relate to other domains such as school-level practices, like professional collaboration and student-teacher relationships, as well as teacher attitudes such as job satisfaction and self-efficacy.

Using the 2009 Assessment Framework of the Programme for International Student Assessment (PISA), learning practices are internationally operationalised to include in-class strategies such as collaborative or competitive peer learning, study strategies, amount of instructional and study time, additional classes outside of

school hours and extra-curricular activities as well as motivation and future plans (OECD 2009b).

1.3 Theory, policy and practice background

This section describes background issues pertinent to this review: the concept of evidence-based policy-making, the dramatic rise of assessment programmes within developing countries, and the different uses of assessment to serve as evidence.

1.3.1 Evidence-based policy-making

Although it is difficult to find an agreed-upon and clear definition of evidence-based policy-making (Marston and Watts 2003), there is a general understanding that the approach involves the ‘rational, rigorous and systematic’ (Sutcliffe and Court 2005) analysis of the best available evidence to inform policy decisions. The development of methods for collating and synthesising research, including systematic reviews such as this one, fall within this framework.

The concept of evidence-based policy-making originated from that of ‘evidence-based practice’ in the health sector, which was itself preceded by the concept of ‘evidence-based medicine’ (Sackett et al. 1996, Sutcliffe and Court 2005). This approach then permeated other policy sectors and fields of practice, including education, social work and criminal justice (Solesbury 2001). The approach and the term were most prominently adopted by the UK government in the late 1990s (Sutcliffe and Court 2005).

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

Table 1.1 Components of the policy cycle and evidence issues

Stage of the policy cycle	Description	Evidence issues
Agenda setting	Awareness and priority given to an issue	The evidence needs relate to identifying new problems or the build-up of evidence regarding the magnitude of a problem so that relevant policy actors are aware that the problem is indeed important. Key factors are the credibility of evidence and also the way evidence is communicated.
Policy formulation	There are two key stages to the policy formulation process: determining the policy options and then selecting the preferred option (see Young and Quinn 2002, pp. 13-14).	For both stages, policy-makers should ideally ensure that their understanding of the specific situation and the different options is as detailed and comprehensive as possible - only then can they make informed decisions about which policy to go ahead and implement. This includes understanding the instrumental links between an activity and an outcome as well as the expected cost and impact of an intervention. The quantity and credibility of the evidence are important.
Policy implementation	Actual practical activities	The focus is on operational evidence to improve the effectiveness of initiatives. This can include analytical work as well as systematic learning around technical skills, expert knowledge and practical experience. Action research and pilot projects are often important. The key is that the evidence is practically relevant across different contexts.
Monitoring and policy evaluation	Monitoring and assessing the process and impact of an intervention	The first goal is to develop monitoring mechanisms. Thereafter, according to Young and Quinn (2002), 'a comprehensive evaluation procedure is essential in determining the effectiveness of the implemented policy and in providing the basis for future decision-making'. In the processes of monitoring and evaluation, it is important to ensure not only that the evidence is objective, thorough and relevant, but also that it is then communicated successfully into the continuing policy process.
No impact	Explicitly no impact on any stage of the policy process	There is an explicit acknowledgement that no evidence from the assessment was used in policy-making.

Source: Sutcliffe and Court (2005, adapted from Pollard and Court 2005).

The most common criticism of evidence-based policy-making relates to its approach to the selection of evidence, specifically the perception of a hierarchy in the way different types of evidence are valued (Sutcliffe and Court 2005, Marston and Watts 2003, Wiseman 2010). This review did not directly engage with this issue as it was concerned with only one specific type of data - that from system-level assessments. The way that data were analysed, however, was still relevant and this formed one of the sub-questions to this review.

Other concerns with the practice of evidence-based policy-making relate to *how* and *when* evidence is considered in the policy cycle. The notion of understanding 'what works' as the central concern of evidence-based policy can limit its utilisation to evaluative-type research, as well as limit its use during policy and programme design (Pawson 2002). However, to truly address the question of 'what works for whom in what circumstances', an evidence base is needed 'in all stages of the policy cycle - in shaping agendas, in defining issues, in identifying options, in making choices of action, in delivering them and in monitoring their impact and outcomes.' (Solesbury 2001, p. 8)

Different issues in the use of evidence are relevant at different stages of the policy cycle. Sutcliffe and Court (2005) outlined these different issues in a table that has been reproduced as Table 1.1.

Table 1.1 served as a synthesising framework in this review to examine the policy cycle stages in which assessment data are utilised as evidence. This framework allowed us to further examine the ways in which assessment data that are utilised in the policy cycle relate to the outlined evidence issues.

1.3.2 Assessment programmes in developing countries

Within developing countries, some aspects of the cultural, economic and political context may represent a considerable challenge to the application of evidence-based policy (Sutcliffe and Court 2005), as well as technical and infrastructural aspects. Factors such as academic and media freedom, the role of civil society, and the stability and openness of political systems are important elements in allowing evidence to be gathered, assessed and communicated to influence policy-making (Sutcliffe and Court 2005). Challenges to these elements are increasingly being overcome, leading to a greater focus on evidence-based policy processes in developing countries (Sutcliffe and Court 2005). This is one possible explanation for the dramatic increase in the implementation of assessment programmes in developing countries.

There has been a documented global rise in the number of countries undertaking national learning assessments (Benavot and Tanner 2007, UNESCO 2008), as well as international and regional learning assessments (Kamens and McNeely 2009). Much of this growth, especially in national learning assessments, has occurred in economically developing countries (Postlethwaite and Kellaghan 2008).

Developing countries only began conducting national assessment programmes in the 1990s (Postlethwaite and Kellaghan 2008), but a global survey found that by 2006 half of all developing countries had carried out a national assessment programme (Benavot and Tanner 2007). In a more recent survey of 151 developing countries (Kamens and Benavot 2011), over two-thirds had participated in at least one international, regional or national assessment in the 1960-2008 period. In fact, 16 countries were found to have participated in all three types of assessments (Kamens and Benavot 2011).

From 1960 to 2008, the majority of developing countries shifted from participating in international assessments to conducting national assessments. However, there

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

was overall growth in national, regional and international assessment programme participation by developing countries (Kamens and Benavot 2011).

By 2008, national assessments constituted slightly under two-thirds of all assessment programmes undertaken by developing countries. Descriptive results from the study by Kamens and Benavot (2011) suggest possible regional differentiation in country participation for the three assessment programme types. Considering participation in international assessments throughout 1960-2008, developing countries in Central and Eastern Europe, Central Asia, the Arab States, Latin American and the Caribbean had much higher participation rates than developing countries in Asia and Africa - South of the Sahara. Examining regional assessments by regional participation, the relationship reverses, and developing countries in Africa - South of the Sahara demonstrate the highest participation rates, followed by developing countries in Latin America and the Caribbean. Considering participation in national assessments, countries in Latin America and the Caribbean have the highest participation rates, and all other regions have participation rates of about 50%, except for developing countries in Eastern Europe and Central Asia.

The increased participation of developing countries in large-scale assessment programmes coincides with a shift in global focus from educational provision such as enrolment rates to improving the quality of education (Braslavsky 2005). Furthermore, there is a growing emphasis on the concept of the development of human capital, as measured by learning assessments, being related to a country's economic growth (Hanushek and Kimko 2000).

The above discussion illustrates that there is a descriptive portrait of developing-country participation in assessment programmes, and a theorised framework to explain policy-making. Still, little is known about the effects of such assessment programmes on education policy and practices in developing countries, and it is therefore a question that this review sought to address.

1.3.3 Assessment data as evidence in education policy planning

Evidence-based education policy-making has been adopted around the world, with Wiseman (2010, p.2) stating that it is 'the most frequently reported method used by politicians and policymakers'. This movement, among others, has provided support for an emphasis on the use of student assessment data in the policy process (Campbell and Levin 2009).

Assessments themselves are complex and can be contested political terrain, serving a multitude of functions which are difficult to reconcile in a single assessment process (Berry and Adamson 2011). There appears to be common understanding, however, that one of the main aims of conducting national assessments or participating in international assessments is to provide information on a country's educational outcomes, which in turn assists policy-makers and other stakeholders in the education system in making policy and resourcing decisions for improvement (Benavot and Tanner 2007, Braun and Kanjee 2006, Forster 2001, Postlethwaite and Kellaghan 2008).

Data resulting from assessment programmes can report the extent to which an education system is teaching its students what is expected, differences in achievement levels by subgroups (such as gender or region) and, if background data are collected, factors that contribute to reaching different levels of achievement. Reporting can be done either as part of the official reporting process or from secondary analysis (e.g. Lietz et al. 2008); this information is often also used to formulate recommendations to improve educational outcomes.

In discussing the uses of evidence for education policy-making, Wiseman (2010) outlined three main goals for evidence-based policy-making: measuring and ensuring quality, ensuring equity, and control, which are referred to throughout this review as accountability. Berry and Adamson (2011) discussed the disparate goals of assessment as diagnosis of strengths and weaknesses in learning, competitive selection and external accountability. Expanding on the notion of 'systemic validity', Braun and Kanjee (2006, p.6) posited that an assessment practice and system is systemically valid if it generates useful information that supports the continuous improvement in access, quality, efficiency or equity within the education system, 'without causing undue deterioration in other aspects or at other levels'.

Combining these concepts for this review, we anticipated finding examples of the use of assessment data as evidence in policy-making in developing countries to fall into the following main groups:

- As a measure of quality level, to diagnose the strengths and weaknesses of a system. The analysis was likely to be achievement-level analysis, comparison between subsections in the assessment content, and trend analysis noting changes in achievement level over time. This type of use will most likely be present during the agenda-setting and policy-formulation stages of the policy cycle, and in impacting policy types such as standard-setting, relative weight of different components of the education sector (e.g. vocational versus academic education) and decisions on system-wide curriculum content.
- To measure and ensure equity within the system. This was likely to rely on analysis that provides comparisons between groups (schools, regions, socio-economic groups) in the system. This use will most likely be present during the policy formulation and implementation stages. Examples of this use include basing the allocation of resources such as funding on assessment results, using information from assessments to design and target interventions to particular disadvantaged or vulnerable groups, as well as the use of assessments as selection criteria for graduation.
- As an accountability tool, or as evidence to practise control over the system. This can cover both internal and external accountability. Accountability with those within a system may include the use of assessment by schools to report to their stakeholders, including the government, as well as policy-makers using changes in assessment results to monitor the outcome of their interventions. We anticipated external accountability to also be pertinent for developing countries, considering the role of international agencies. This can take the form of formal reporting requirements as well as less direct influences. This use will most likely take place during the monitoring and evaluation stage of the policy process.

From the application of our initial conceptual framework to the literature, an additional use of assessment data as evidence in policy-making was incorporated into the conceptual framework:

- As evidence to apply leverage to and prioritise pre-existing political agendas and policies within the policy-making process. This goal is often highlighted in critiques of assessment programmes. Leverage is distinct from accountability in that it does not have the aim of practising control over the system in order to eventually improve educational outcomes, but is intended to promote policies that are aligned with a political agenda.

For all the above uses, it is important to note that the effect of any analysis or research may not always be positive, or supportive of certain propositions. During

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

the formulation of a policy, assessment outcomes may demonstrate difficulties in a certain intervention, rather than feasibility (Husén and Kogan 1984). They may also show negative outcomes of a policy during evaluation.

In countries where the utilisation of assessments in the policy-making process is explicit, controversies are rife around the use of assessments as the centrepieces of policies. Examples of this include the use of assessments in publically comparing schools and rewarding teachers in Australia (Bantick 2011, Hardy and Boyle 2011, Topsfield, 2011) and, in the USA, the over-emphasis on assessments as the basis of accountability systems, including the use of assessments to close down schools whose students fall below state proficiency standards (Darling-Hammond 2004, Ravitch 2010).

Little has been written about the optimal use of assessment findings or the effects of basing policy decisions on the findings in developing countries (Kellaghan et al. 2009). As little is known even about how assessments are used in policy-making in developing countries (Kellaghan et al. 2009), it is not surprising that Kamens and Benavot (2011, p. 296) concluded that 'how countries conduct and use assessments, and the policies surrounding these uses, are ripe subjects for comparative research'.

1.3.4 Potential facilitators and barriers to the utilisation of assessment data

This review also sought to collate and synthesise evidence on the facilitators and barriers to the use of assessment data to inform policy-making in developing countries. A number of such possible factors have been described in the literature on this topic. These can be related to the nature of the assessment programme itself, the analysis of assessment outcomes, the dissemination of findings from the programme, the nature of the education system and the nature of the political system and wider context.

Factors that relate to the nature of the assessment include:

- The soundness and appropriateness of the assessment instrument, sampling approach and administration procedures (Braun and Kanjee 2006, Kellaghan et al. 2009);
- How well the assessment programme is integrated into existing structures, policy and decision-making processes (Kellaghan et al. 2009); and,
- The level of involvement of policy-makers in the design and implementation of the assessment programme (Kellaghan et al. 2009).

Factors that relate to the analysis of assessment outcomes include:

- Whether secondary or in-depth analysis of data is undertaken, beyond initial descriptions (Wiseman 2010); and
- Whether analysis is undertaken with a focus on diagnosing issues in the education system, including identifying factors associated with high and low achievement (Kellaghan et al. 2009).

Factors that relate to the dissemination of findings and analysis from the assessment programme include:

- The timeliness of results dissemination (Kellaghan et al. 2009);
- The extent to which key users receive appropriate reports of findings from the programme, including senior policy-makers, curriculum developers, teachers and the media (Kellaghan et al. 2009, Postlethwaite 1984);
- Whether the assessment findings are communicated in a way that is appropriate to the needs of, and can be understood by, policy-makers and other stakeholders (Kellaghan et al. 2009, Postlethwaite 1984);

- The level that policy-makers are able to understand the findings and critically appraise them (Campbell and Levin 2009, Davies 2004 in Sutcliffe and Court 2005, Postlethwaite 1984); and,
- The value that is placed on assessment findings, as well as research inputs and evidence in general, by policy-makers (Campbell and Levin 2009, Marston and Watts 2003).

Factors that relate to the nature of the education system include:

- The effectiveness with which the education system functions (Braun and Kanjee 2006);
- The strength of teachers' unions and their role in policy-making (Kellaghan et al. 2009); and,
- Whether there are good communication channels or distribution systems from the decision-making and research stakeholders to schools (Postlethwaite 1984).

Factors that relate to the nature of the wider political system include:

- Whether there are political sensitivities to making findings public (Kellaghan et al. 2009);
- The level of decentralisation and openness of the political system (Kellaghan et al. 2009, Sutcliffe and Court 2005);
- The level of public representation and strengths of structures for aggregating and arbitrating interests in society (Kellaghan et al. 2009, Sutcliffe and Court 2005);
- The extent of academic and media freedom and the strengths of civil society (Sutcliffe and Court 2005);
- The existence of conflict or political volatility (Sutcliffe and Court 2005);
- The strength of accountability systems (Sutcliffe and Court 2005); and,
- The role of external (multilateral and bilateral) agencies in the system (Kellaghan et al. 2009).

1.4 Research background

The extent to which national assessment findings are having an impact on policies and resource allocation decisions, or on teaching and learning processes, has come under scrutiny in recent years. Despite the growing popularity of national assessment programmes and the potential value of the information they can provide, a 2009 World Bank report found that available evidence indicates that findings of these programmes are not widely used (Kellaghan et al. 2009). In drawing this conclusion, the authors noted that limited information was available on this topic (Kellaghan et al. 2009).

Reviews have, however, been conducted on the impact of the International Association for the Evaluation of Educational Achievement's (IEA's) Trends in Mathematics and Science Study (TIMSS) on policies in developing countries, some of which were undertaken to monitor the effectiveness of the support the World Bank provided to these countries to enable them to participate. Reviews have also been conducted as part of the accountability process in regional assessment programmes such as the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), with participating ministries periodically required to report any observed impacts of the assessment on their countries' policies.

A 2009 OECD evaluation of the policy impact of PISA on participating countries and economies found that while the influence of PISA on policy formation is increasing over time at both the national and local levels, the policy impact of PISA is greater at the national level than the local level. The evaluation report identified that policy-makers are the most important stakeholders in PISA participation and

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

results, rather than other stakeholders such as local officials and school principals (OECD 2009c), which could be a factor in the smaller impact of assessments undertaken at the local level.

1.4.1 Systematic reviews

The EPPI-Centre's evidence library houses six systematic reviews on the topic of assessment. These reviews examined the impact of different assessment forms on teaching and learning - in other words, they were focused on classroom-level impact. None of the reviews looked at the impact of assessments at the level of the education system or, as was the case for this review, on the policy-making process.

To the best of the authors' knowledge, no systematic review has been completed on this topic, much less within the context of developing countries.

1.5 Authors, funders, and other users of the review

1.5.1 Funders of the review

This review was funded by AusAID. It received one of 32 awards for systematic reviews to investigate the impact of development interventions under a joint call for proposals between AusAID, the UK's Department for International Development (DFID) and the International Initiative for Impact Evaluation (3ie). The scheme was designed to strengthen the international community's capacity for evidence-based policy-making.

In cooperation with partner governments and other development agencies, AusAID has supported the development and improvement of systems for monitoring learning outcomes in its partner countries. AusAID's interest in understanding the impact of these interventions led to their submission of the original question (later revised - see below) as the first in the joint call for proposals:

What are the impacts of standardised national assessments (for example, of reading and numeracy) on policy, resource allocation and learning achievement for primary and secondary schools in developing countries?

1.5.2 Authors of the review

The review was undertaken by a team of researchers from the Australian Council for Educational Research (ACER), led by Dr Petra Lietz. At the institutional level, ACER has extensive experience working with large-scale assessment programmes internationally. The ACER authors brought together expertise in international assessments, in working with policy-makers in developing countries, and in undertaking literature searches, information retrieval, as well as in undertaking reviews and syntheses. Additionally, the team was supplemented by a systematic review expert from the Joanna Briggs Institute (JBI) at the University of Adelaide, Australia, and received support from the EPPI-Centre at the University of London, UK. Please see Appendix 1.1 for detailed information of the research team.

1.5.3 Peer review and advisory groups

Appear review group was hosted by the EPPI-Centre, with whom this review was registered. After providing feedback on the title registration, EPPI-Centre also coordinated the review of the Protocol document and the draft report for this review.

The peer review group included relevant researchers, and policy-makers nominated from the organisations funding, advising and undertaking the review. These peer reviewers had relevant systematic review, policy and topic-area interest expertise.

Peer review group members included:

Peer review expertise	Peer review member	Affiliation
Systematic review	Nominated by EPPI-Centre	
Policy area	Nominated by AusAID	
Topic area interest	David Rutkowski, PhD	Assistant research scientist, Center for Evaluation and Education Policy, Indiana University, USA

An advisory group provided further feedback and guidance regarding the development of the review, which included the conceptual framework, identification of relevant literature, inclusion and exclusion criteria, and review and synthesis of the evidence. This group provided specific feedback relevant to members' expertise, which included pertinent theoretical, policy, and practitioner knowledge of assessment programmes in developing countries. Furthermore, the advisory group members were also able to provide a critical voice and knowledge of stakeholders from developing countries to better inform the relevance of this review.

Advisory group members included:

Name	Affiliation
Lucrecia Santibanez, PhD	Economist, Center for Latin American Social Policy, RAND Corporation
Maurice Robson, PhD	Chief of Education Section, Pakistan, United Nations Children's Fund

1.6 Review questions

As outlined in earlier sections, this review aimed to examine the body of evidence on the impact of large-scale assessment programmes on educational policies, particularly policies regarding resource allocation and teaching and learning practices in developing countries. Due to concerns with the availability of literature and methodological issues, the reference to impact on learning achievement in the original question was removed and amended to impact on teaching and learning practices. Teaching and learning practices act as the mediating factor in the mechanism that allows assessment programmes to impact upon learning outcomes.

Furthermore, reference to resource allocation and teaching and learning practices was clarified to mean educational policies that specifically concern themselves with those domains, rather than an evaluation of policy implementation regarding resource allocation and teaching and learning practices.

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

The main question addressed by this review is therefore as follows:

What is the evidence on the impact of large-scale (i.e. national and international) assessment programmes on education policy-making, particularly policies regarding resource allocation and teaching and learning practices in developing countries?

It also sought to address the following sub-questions:

- i. What are the characteristics of the large-scale (i.e. national and international) assessment programmes that have informed education policy-making in developing countries?
- ii. How are the data from assessment programmes used to inform education policy-making in developing countries?
- iii. At what stages of the policy process are data from assessment programmes being used in developing countries?
- iv. What educational policies in developing countries have resulted from the use of assessment data?
- v. What are the facilitators and barriers to the use of assessment data in education policy-making?

It was anticipated that there would not be a significant body of literature that directly addressed these questions. Therefore, this review utilised approaches that allowed us to accommodate a broad range of literature and from it synthesise aspects that were relevant to answering the review questions.

1.7 Systematic review background

Systematic reviews are considered to be the highest level of evidence generation, as they aim to systematically locate all of the evidence about a question, appraise it, extract data and present a summary of the findings. Systematic reviews place emphasis on comprehensive searching, meticulous screening and critical appraisal (against pre-determined criteria) of the methodological quality of the included papers (JBI 2011a). Systematic reviews are important tools that help clinicians, researchers and policy-makers alike to summarise the existing information in order to make evidence-based decisions. Systematic reviews are different from narrative or literature reviews, because they are based on protocol, hypothesis and selection criteria prepared before conducting the review and seek to preserve objective examination of the data before inferences are drawn.

The methodology of systematic review is now entrenched as the gold standard of scientific inquiry and reporting to inform policy, practice and guideline development (JBI 2011a). Many countries now require a systematic review to demonstrate a gap in knowledge before competitive funding will be allocated to new primary research and this is just one indication of the ascension of systematic reviews (Tricco et al. 2008). Systematic reviews have replaced the ubiquitous randomised controlled trial as the top tier of evidence in most hierarchies or evidence tables, particularly within the health sciences, but more broadly, this is also the case across the social sciences.

A systematic review is a piece of scientific research that uses existing literature (published and unpublished) as its source of data. Systematic reviewers develop an *a priori* protocol to guide the conduct of their review. This, along with other features of the systematic review, contribute to its scientific validity and its global recognition as the ideal basis from which to inform policy and practice related issues (JBI 2011a). The protocol is operationalised in the same way as a primary research protocol. It describes the review question and sets the parameters that the review will follow, with the *a priori* nature of a review protocol avoiding the

risk of generating results that have been led by the reviewer rather than independently and transparently extracted from studies. The protocol also includes descriptions of the types of participants and participant characteristics, the intervention or phenomena of interest and what the comparator intervention is, as well as a list of primary and (less often) secondary outcomes of interest. The review protocol also describes the methods by which the validity and reliability of studies will be assessed, how data extraction will be undertaken and what methods of synthesis will be used for the types of outcome data that are extracted (JBI 2011a).

The science of systematic reviews evolved within the positivist paradigm and while aspects and fine detail may be debated there is broad consensus that a systematic review can be identified by a particular set of characteristics. These, as Tricco et al. (2008) suggest in their analysis of published systematic reviews, tend to focus on minimising the risk of bias in the following domains:

- The development of and adherence to an *a priori* protocol to reduce risk of researcher influence particularly in relation to the results (performance bias);
- Methods for the identification of literature to be assessed for inclusion (publication and citation bias);
- Methods for how studies are selected for retrieval (selection bias); and
- How the quality of identified studies is rated or appraised, leading to a decision on whether they should be included or not (risk of assessment bias). (Tricco et al. 2008)

These accepted conventions sit well within the positivist paradigm as they are objective measures with known impact on reducing the risk of bias. Crotty (1998) identified these distinctions in his foundational text on research in the social sciences by highlighting that the attributes of positivism are associated with objectivity. What we study from this perspective has meaning of its own and this meaning can be understood if our methods ensure the researcher and the researched do not cross-contaminate; if they use empirical methods of measurement; and if the line of inquiry is one that seeks to discover meaning rather than ascribe meaning (Crotty 1998). In this way, Crotty draws out the distinguishing features of quantitative research and the focus on objectivity that forms a useful point of reference for consideration of subsequent developments in the conduct of systematic reviews in the critical and interpretive paradigms. While this acts to situate a reference point, there is no implication that one paradigm or methodology is somehow inferior or less empirical than another. Therefore, the strength of systematic review methodology is not reliant upon whether the focus is on quantitative analysis, qualitative analysis, or any other particular method of synthesis; it is, in fact, reliant upon the development of a rigorous protocol that outlines detailed and auditable strategies for each stage of the review, where these are adhered to by the review team.

1.7.1 Framework synthesis approach

The selection of the framework synthesis approach to undertake this review was made based on an understanding of the nature of available literature. Despite the increase in developing countries' implementation of and involvement in large-scale assessment programmes, very little was known about their impact, and how they were being used by policy-makers and practitioners. Our own initial exploration of the literature on this topic led to a similar early conclusion. We anticipated that the literature on this topic and in this setting would be overwhelmingly qualitative in nature. Furthermore, we anticipated that narrative and other textual papers would form a significant part of the literature. Taking into account this anticipated nature of the literature and the understanding that this topic was in general still

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

under-researched, the use of an iterative process like the framework synthesis approach seemed most appropriate to accommodate varied literature.

This approach utilised an *a priori* conceptual framework that guided the extraction and synthesis of findings (Barnett-Page and Thomas 2009). The use of an *a priori* framework may take into account existing research and policy issues (Carroll et al. 2011) as well as prior experience and knowledge (Oliver et al. 2008). The use, therefore, of a deductive process, which uses an *a priori* framework, was an optimal approach to address applied policy questions (Dixon-Woods 2011). Furthermore, this approach fulfilled a pragmatic imperative by enabling researchers to search and synthesise the evidence from a large volume of literature to address timely policy issues (Dixon-Woods 2011). As there was a body of literature dealing with the types of country participation in assessment programmes, and a theoretical body of literature dealing with education policy cycles and the use of evidence-based policy-making, this approach used pre-existing models to enable the coding and synthesis of the included studies (Carroll et al. 2011) to examine the impact of national and international assessment programmes on education policy-making and practices in developing countries.

The research team had developed an initial conceptual framework which incorporated the simplified model of the policy cycle into a larger framework of assessment programme impact on education policy-making. This initial conceptual framework was used to identify, appraise and analyse relevant literature as well as to code and synthesise it (Oliver et al. 2008).

This framework synthesis approach shifted to an iterative process in the later stages as *de novo* topics emerged from the data (Dixon-Woods 2011). These *de novo* themes were incorporated into the pre-existing conceptual framework. The analysis of the data thus became an iterative process as *de novo* themes emerged and the initial conceptual framework was modified. Therefore, the use of an initial conceptual framework, and the later development and incorporation of new themes and topics into the framework, shifted the framework synthesis approach from a deductive to a more inductive analysis of the data. (See the conceptual framework in Figure 1.2.)

1.7.2 Focus of review

The population of interest for this review is education policy-makers and practitioners in developing countries. The review sought to explain the impact of assessment programmes on the policy-making process and practices. The actions of these groups are what lead to changes in educational policies and practices in their countries. This review, therefore, concerned itself with how assessment programmes influence these actors.

Figure 1.2 visually depicts the *a priori* conceptual framework that was developed for this review.

The interventions that this review was concerned with were assessment programmes, more specifically large-scale standardised assessment programmes as defined in Section 1.2 above. These may include national, regional and international programmes. The conceptual framework presupposes that data from these three types of assessment programmes are analysed in certain ways, and that the ways in which the data are analysed may impact upon how assessment findings are used by policy-makers. Concerning assessment programmes, the review considered sub-questions one and two:

- i. What are the characteristics of the large-scale assessment (i.e. national and international) programmes that have informed education policy-making in developing countries?

- ii. How are the data from assessment programmes used to inform education policy-making in developing countries?

Although the review was not designed at the outset to make comparisons between different types of assessment programmes, some potential comparisons that could have arisen from the findings included those between population and sample-based assessment programmes, and between high-stakes and low-stakes assessment programmes.

The conceptual framework outlined the context, or goals of evidence-based policy-making: quality, equity and accountability, as described in section 1.3.3. What the review examined as its outcome of interest were changes in the education policy-making process (including at the policy development, implementation and evaluation stages) and in the teaching and learning practices in developing countries. As such, the conceptual framework considered the goals of evidence-based policy-making and the stages of the policy process and associated issues in the use of evidence (see Table 1.1) in examining changes in policy-making processes, and particularly policies regarding teaching and learning practices. Concerning policy-makers and the policy process, the review considered sub-questions three and four:

- iii. At what stages of the policy process are data from assessment programmes being used in developing countries?
- iv. What educational policies in developing countries have resulted from the use of assessment data?

Lastly, the conceptual framework aimed to synthesise evidence to address review sub-question five:

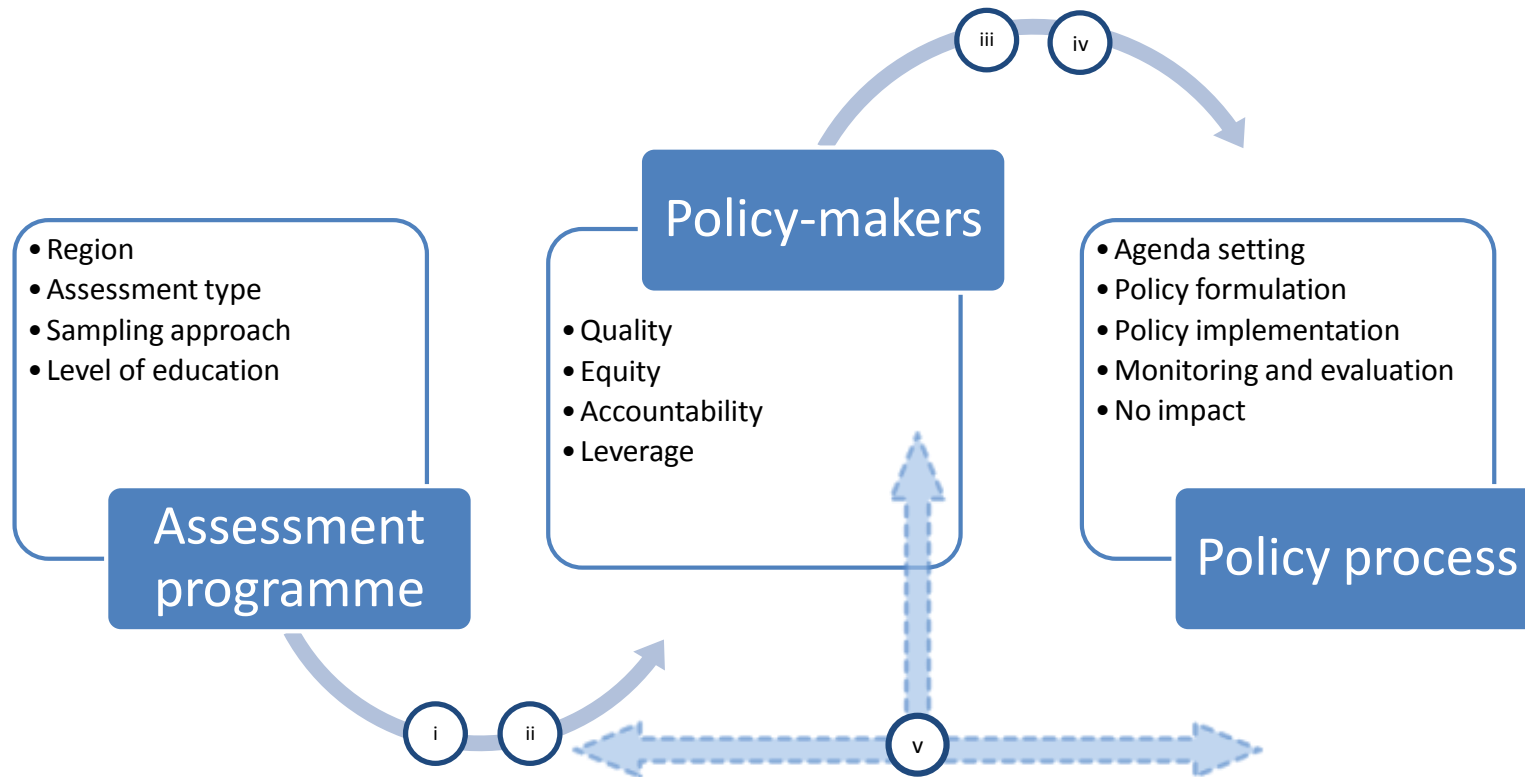
- v. What are the facilitators and barriers to the use of assessment data in education policy-making?

Considering facilitators and barriers, factors may relate to the assessment programme and the analysis of results, the policy-making context and the stages in the policy cycle. These related factors, which were considered in the conceptual framework, are described in detail in section 1.3.4.

The complex relationship between assessment programmes and these processes cannot be easily captured in the experimental or quasi-experimental impact evaluation designs that are commonly the focus of systematic reviews. This led to the selection of a framework synthesis approach to this review.

Study designs that were likely to address this topic and could be captured under this approach included policy analysis papers, reports on the evaluation of particular assessment programmes that described the uses of assessment outputs and/or the impact of the assessment process, and textual papers, incorporating narrative and expert opinions.

Figure 1.2 Conceptual framework



- i. What are the characteristics of the large-scale (i.e. national and international) assessment programmes that have informed education policy-making in developing countries?
- ii. How are the data from assessment programmes used to inform education policy-making in developing countries?
- iii. At what stages of the policy process are data from assessment programmes being used in developing countries?
- iv. What educational policies in developing countries have resulted from the use of assessment data?
- v. What are the facilitators and barriers to the use of assessment data in education policy-making?

2. Methods used in the review

Outline of chapter

This chapter outlines the methodologies that were employed for the review. The first section describes the most immediate as well as the anticipated users of this review and how they were involved in the review processes. This is followed by the strategies for searching, including (and excluding) and reviewing the literature that was used.

2.1 User involvement

The most immediate users of this review are those within the funding body, AusAID, who are directly or indirectly involved in the funding and management of AusAID's involvement with systems for monitoring learning outcomes in its partner countries. This includes staff within the Education Thematic Group, Education Advisors based both in Canberra and in country offices, as well as staff and consultants working with partner governments that are considering developing, amending or becoming involved in assessment programmes.

AusAID formed a consultative group comprised of these relevant personnel, and they played a key role in providing direction to the authors during the inception stage of the review. They were also consulted during review stages, a process coordinated by the EPPI-Centre.

It was anticipated that this review would be of use to other agencies that are considering or evaluating support to assessment programmes. These agencies may include government bodies - particularly in developing countries - that are considering an assessment programme or seeking to evaluate the way they make use of existing programmes. Additionally, agencies that are involved in the design and implementation of assessment programmes may be interested in how these programmes are being used by policy-makers and practitioners.

Furthermore, this review would be of interest to academics, researchers, outside organisations and accountability bodies that have interests in disseminating and communicating the results of assessment programmes to policy-makers and practitioners, in order to inform evidence-based policy-making and practice.

The anticipated users were a potential source of materials for inclusion in this study. We therefore involved these groups from an early stage in the review process by contacting them to inform them of the review and ask for leads to potentially relevant literature. They were identified through the authors' extensive network of stakeholders in assessment and education policy-making in developing countries, and included:

- AusAID's Education Advisors and Education Specialists network;
- National Study Centre representatives from developing countries for IEA and OECD assessment programmes;
- Contacts within regional assessment programmes in developing countries;
- Authors' contacts within ministries of education in developing countries; and,
- Authors' contacts with academics in developing countries.

The Principal Investigator presented the preliminary findings of the review at a conference in Bangladesh of health and education policy advisers with responsibilities in South-east Asia. In addition to policy advisers, the conference was attended by stakeholders from many international development agencies, in-

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

country health and education specialists, representatives of donor agencies and other education stakeholders. The presentation highlighted the importance and usefulness of the systematic review methodology to address pressing policy concerns that often face policy-makers and development stakeholders in developing countries. Feedback to the review revolved around the rigour of the systematic review process compared with traditional literature reviews and cross-checks regarding assessment programmes that were perceived to be influential in education policy-making. The participants were also used to request leads for grey literature relevant to the review.

Members of the identified groups were included in the distribution list for the draft report emanating from the review as well as in the final dissemination of findings.

2.2 Identifying and describing studies

This section describes the approach that was undertaken to identify the literature that was included in this review.

2.2.1 Defining relevant studies: inclusion and exclusion criteria

Based on an initial exploration of the literature, it was anticipated that there would not be a significant body of literature that directly addressed the review main question and sub-questions. We also considered that the methods for answering these questions were not limited to particular study designs. Therefore, we did not exclude studies based on pre-determined study design conditions.

Rather, the inclusion criteria for this review were based on relevance criteria, or how well the literature would be able to answer the review questions. To be included in the review, a report, study or paper needed to meet all of the following criteria:

- It made reference to a sub-national, national, regional or international standardised assessment or testing programme;
- It made reference to at least one developing country; and,
- It made an explicit reference to the link between an assessment programme and (i) a stage in the policy-making process (policy design; evaluation of policy options; policy implementation; policy evaluation), or (ii) a change in policy within the education sector (e.g. allocation of resources in education, curriculum design, standards development), or (iii) a change in teaching or learning practices at the classroom or school level.

From the listed inclusion criteria, a set of exclusion criteria was developed to apply to the title and abstract screening process. As title and abstract screening was double-coded by team members, applied exclusion criteria were more easily able to be reviewed and negotiated in cases of disagreement. The list of exclusion criteria used in the review software is listed in Appendix 2.1.

Furthermore, in addition to basing the inclusion criteria on relevance to the review questions, included literature met additional criteria based upon the theoretical foregrounding in the literature review.

As outlined in section 1.2.1, although assessments may still serve a 'gate-keeping' function where there are limitations in the availability of places in subsequent stages of schooling in some countries, they are mainly used to monitor and evaluate the quality of student learning outcomes (Postlethwaite and Kellaghan 2008) and are designed to enable comparisons over time (Greaney and Kellaghan 2008). As outlined in section 1.3.2, much of the growth in assessment programmes, especially in national learning assessments, has occurred in economically

developing countries (Postlethwaite and Kellaghan 2008), which only began in the 1990s. Therefore an included report, study or paper had to meet the following criteria:

- It made reference to an assessment programme conducted in either primary or secondary education or in both; and,
- It was published or released between 1990 and 2011.

Lastly, as outlined in section 2.2.2 below, searching only for English language literature would increase the possibility of publication and positive biases. We therefore expanded the searches to non-English-language studies. Between them, the review team members had a working knowledge of French, German, Indonesian and Spanish.

2.2.2 Identification of potential studies: search strategy

Our search strategy sought to identify published as well as 'grey' literature. In identifying databases for searching, we purposefully included those that included grey literature. Additionally, we supplemented our searches of databases, journals and the internet with literature sourced by contacting groups and individuals, such as academics.

More specifically, we employed five strategies for identifying potential studies:

- i. *Electronic searches of bibliographic databases:*

Databases available to ACER were utilised, including the Australian Education Index (AEI), Education Resources Information Centre (ERIC), Education Resource Complete (ERC), British Education Index (BEI), Scopus, Eldis, Asia-Studies Full Text Online, Google Scholar and the British Library for Development Studies (BLDS). These initial searches were conducted by experienced information librarians who also set up 'alerts' to capture further material during the review, such as Google Scholar alert links made for 'national student assessment' and 'international student assessment'. Keywords were employed to narrow down the intervention: (national OR international OR "large-scale" OR system-wide OR standardised OR standardized OR standard) AND (assessment* OR exam* OR test*) OR "learning outcomes" OR ((student OR learning) AND achievement). Terms were also employed for controlled-vocabulary searches using database specific keywords: ("Student evaluation" OR "Student assessment" OR "Educational testing" OR "Achievement tests" OR "Standardized tests" OR "Standardised tests" OR "Testing programs" OR "Testing programmes" OR "National standards" OR "National competency tests" OR "Testing" OR "Educational tests and measurements" OR "High-stakes tests" OR "Academic achievement testing" OR "Academic achievement" OR "Competency based educational tests" OR "Examinations" OR "National competency based educational tests" OR "Student Assessment" OR "National standards").

These were combined with free-text terms to describe the intervention, for example, (National exam*). They were also combined with controlled-vocabulary terms to narrow down the specified level of education in which the intervention was conducted: ("Elementary education" OR "Education elementary" OR "Elementary secondary education" OR "Secondary education" OR "Education secondary" OR "Primary Education" OR "Middle school education" OR "Primary secondary education").

The research team conducted several test searches to assess if the defined search terms were appropriate to locate relevant literature. Initially, it was

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

proposed to include search terms to describe types of impacts of assessment programmes: (“impact” OR “effect” OR “use” OR “utilisation” OR “utilization” OR “benefit” OR “consequence”). For databases that were not focused on development or developing countries, we supplemented these with keywords that focused the search further, such as: (“developing” OR “third world” OR “impoverished”) AND (“nation” OR “country” OR “region”).

Using terms to describe types of impacts was found to be too restrictive, and it was decided to exclude these terms from the electronic database searching in order to increase the number of relevant records returned. In addition, we used the names of specific developing nations as identified by AusAID, the International Monetary Fund (IMF) and the World Bank, such as Indonesia, Mexico, etc., and the names of regions, such as Africa and Latin America. Where there was a geographic descriptor field in the record, we searched that field for the names of countries or regions; otherwise we used the subject or identifier fields, where those existed. For those databases not focused on education, we added the search term ‘education’. The above strategies and search terms were modified to suit each database. Appendix 2.2 documents all descriptors and keywords used for each database to permit replication.

ii. *Targeted searches of contents pages of key journals*

Targeted searches were conducted to overcome the delay between dates of publication and appearance in bibliographic databases. These included *Assessment in Education: Principles, Policy and Practice*; *Educational Research*; *Evaluation and Research in Education: Asia Pacific Journal of Education*; and *International Journal of Educational Development*. We identified key regional and national journals on assessment and education policy.

iii. *Targeted searches of online holdings of international/regional agencies, research firms and national ministries.*

This included targeting known international, regional and national assessments in the repositories of agencies that manage international and regional assessment programmes, such as IEA, OECD, SACMEQ and the Latin American Laboratory for Assessment of the Quality of Education (LLECE), as well as agencies that provide support and research into assessment programmes in developing countries, such as the World Bank, DFID, AusAID and UNESCO. Additionally, this included the publications of relevant research bodies, such as the Research Triangle Institute (RTI), Educational Testing Service (ETS), Cambridge Assessment, Consortium for Research on Educational Access, Transitions and Equity (CREATE), the Association for the Development of Education in Africa’s ADEAnet, Campbell Collaboration, and 3ie research bodies of key national ministries.

iv. *Citation chasing*

This involved checking the references of relevant publications to identify possibly relevant literature as well as forward-citation tracking using Scopus, or searching through the list of papers/studies that had cited relevant literature.

v. *Contacting relevant groups and researchers in this area*

This was an iterative strategy, where query emails for potentially relevant literature were sent to possible sources, such as ACER (India) and PISA national centres, as well as telephoning and emailing key authors identified through the literature search. A ‘snowballing’ technique was employed, which meant asking contacts to refer us to other researchers or authors who had access to additional literature.

As limiting the searches to the English language increased the possibility of positive and publication biases we expanded the searches to non-English studies, and used search engines such as Google Scholar. The research team included those with a working knowledge in French, German, Indonesian and Spanish. We believed that this covered major languages of publication in the given setting. These searches were further undertaken in non-English, regional databases, such as Latin American Journals Online (LAMJOL), and regional databases such as African Journals Online (AJOL), which offer peer-reviewed articles from Southern scholars.

A database system was set up to keep track of, and to code, studies found during the review. Titles and abstracts were imported directly from online databases or, when necessary, entered manually, into our database. All database searches were conducted between September and December 2011.

The full search strategies for the electronic databases are contained in Appendix 2.2.

2.2.3 Screening studies

Following the search phase, and addition of studies to the EPPI-Reviewer software, all studies were subject to standardised screening. The screening process was undertaken by all members of the review group, and followed a meeting to discuss the criteria and establish how the screening would be undertaken. The approach was based on the standard methods of screening in systematic review methodology (See Appendix 2.1).

Each reviewer screened a subset of papers for relevance to the review by assessing the title and abstract for keywords related to the review inclusion criteria.

Methodology was also examined to ensure papers were a good fit with the review question. This phase of screening was not intended to exclude papers based on quality, and was instead used to select papers considered relevant to the review question. The exclusion criteria for this review were not based upon issues of methodology, but goodness of fit with the review question. Therefore, papers were only excluded on screening if they were not relevant to the review question or if, for example, they were published prior to 1990.

2.2.4 Methodology and rationale for retrieving full-text for screened titles and abstracts

The 363 records selected, based on title and abstract, for further detailed examination required full-text to be located and uploaded into the EPPI-Reviewer software. Where there was EBSCO host accessibility to a journal article, an 'availability' link had been automatically imported with the record.

Further documents were searched for on the internet using title (or part thereof), keyword, author and appropriate operators relevant to the search engine. Many documents were found freely available in PDF or Word format. Records were located on national and international agency websites and a number of articles through access to JSTOR, DOAJ (Directory of Open Access Journals) and SCIELO (Scientific Electronic Library Online).

The Cunningham Library catalogue was searched for journal articles, reports and book chapters and identified material scanned and uploaded into the software. Fourteen articles and three books were sourced and retrieved through inter-library loans.

When necessary, and where contact details were available, authors or agencies were contacted and a copy of the document requested.

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

2.2.5 Quality assurance process for screening and keywording studies

The key to ensuring the auditability and transparency of decision making in systematic reviews is based upon the application of standardised processes, definitions and terms (JBI, 2011a). In order to ensure there was a shared understanding of the processes embedded in EPPI-Reviewer, and that all reviewers were implementing the operational terms consistently, double-coding was undertaken on a subset of the studies.

The screening criteria were based upon the questions in the review, and were therefore sensitive to studies or papers which comprised the materials in the review and that were of most relevance to the review topic. The screening had a particular emphasis on identifying papers that involved standardised national, regional or international evaluations relevant to policy perspectives.

On completion of screening, a follow-up meeting was held by the review group to discuss the process and the results in terms of particular studies that needed further review or consideration, and to cross-check decisions against the review inclusion criteria. This is a common method of validation within systematic reviews to ensure that rigour of screening is maintained.

2.3 In-depth review

2.3.1 Moving from broad characterisation (mapping) to in-depth review

After the inclusion and exclusion criteria were applied to the studies and papers, keywording strategies were tested on a subset of the material. The subset of papers included 20 studies that were read and coded by members of the review group. On completion of coding, a teleconference was held to discuss the process and clarify any potential changes that might be needed based upon this test dataset. Operational definitions had been developed for each code item, and these were also discussed and reviewed to ensure that the review group had a shared understanding of the meanings of each definition and how it should guide the coding process. The review group met and discussed the relationship between the initial conceptual framework and the data that were emerging from the literature. The discussion was focused on establishing the policy-related aspects of each paper, and the methodological quality and categorisation of papers in terms of their goodness of fit with the five review sub-questions. These were:

- i. What are the characteristics of the large-scale (i.e. national and international) assessment programmes that have informed education policy-making in developing countries?
- ii. How are the data from assessment programmes used to inform education policy-making in developing countries?
- iii. At what stages of the policy process are data from assessment programmes being used in developing countries?
- iv. What educational policies in developing countries have resulted from the use of assessment data?
- v. What are the facilitators and barriers to the use of assessment data in education policy-making?

The initial framework was based on core concepts related to the scope of programmes as being sub-national, national, regional, or international. Within this construct, all papers could be coded, and therefore these constructs were found to be robust. The three domains of policy were established as being (i) quality, (ii) equity and (iii) accountability/control. Under this broad schema, sub-domains of agenda setting, policy formulation, policy implementation, and monitoring and policy evaluation were identified, although monitoring and policy evaluation only

related to accountability/control while policy formulation was shared across both quality and equity. It became apparent that the initial framework provided an adequate representation of the scope of literature (with minor changes as described below) although, as data extraction progressed, some adjustments were made. These allowed for greater clarity in distinguishing the types of literature being included, and in particular, papers that reflected an interest in re-presentation or interpretation of findings without addressing aspects of policy were classified as 'academic papers'. The ability to identify aspects of the studies that could be identified as facilitators or barriers to policy at the level of agenda setting, policy formulation, policy implementation, or policy monitoring/evaluation was also clarified with additional criteria created to separate the coding of facilitators and barriers. Separation of the coding for facilitators and barriers enabled a deeper examination of and extraction from the literature.

This process represented a conceptual shift in the theoretical framework as there were changes in how the coding proceeded, what data were extracted and how facilitators and barriers were identified and evaluated in terms of impact on particular facets of policy. The synthesis therefore moved from a deductive approach based on extant findings, to an iterative, interpretive approach as *de novo* themes emerged and were integrated into the code set.

2.3.2 Detailed description of studies in the in-depth review: EPPI-Centre and review-specific data extraction

The EPPI-Centre approach to coding was implemented at two levels. The first level was to indicate which characteristics from the code set were evident in each individual paper. The total pool of papers was divided across pairs in the review group, and subject to detailed analysis and coding. Identifying the key characteristics allowed synthesis based on standardisation of findings extracted from papers, and facilitated integration of findings across papers. The second level of coding consisted of extraction of key data illustrative of each code from the included papers. These extractions were text from different sections of the reports that gave weight to the particular codes. The extraction of key data provided the richness of text to communicate the context and detail explaining why a particular code was selected. Supporting data provided the granular perspective on the relevance and applicability of data to the review questions. (See Appendix 2.3 for the data-extraction tool.)

Two reviewers independently used a standardised tool to extract data from the included studies. Detailed data from studies included the following:

- Aims and rationale of the study;
- Design;
- Description of the sample;
- Data collection methods; and,
- Data analysis methods.

As with the extraction of codes described above, the review group discussed the standardised details extracted in advance and pre-determined definitions were used to inform decision making on what text was representative of the data fields. Data extraction represents a significant aspect of the review process, and the use of pre-identified fields and standardised operational definitions for those fields is crucial to promoting the transparency and auditability expected of systematic reviews. In addition to the above information, the reviewers also applied a standardised quality assessment to the included studies (described in section 2.3.3 below).

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

2.3.3 Assessing quality of studies and evidence for the review question

Studies included in this review had met the requirements of a series of screening criteria (described previously). The first phase of critical appraisal was to identify the nature of the evidence, whether the papers were to be categorised as research (including surveys, case studies, and interview-based or other qualitative or quantitative papers) or non-research/textual papers (including narrative, opinion or expert papers). Based on this grouping, a second-stage analysis was proposed using critical appraisal checklists that were modified from those developed by JBI (JBI 2011a). Two appraisal instruments were considered, one for the evaluation of rigour in text and opinion based on the JBI Narrative, Opinion Text and Review Instrument (JBI 2011b). This instrument had been designed by the institute based on a conceptual approach to identifying validity of non-research-based evidence. The second instrument was from the JBI Qualitative Assessment and Review Instrument (JBI 2011b). This tool focuses on the internal validity of qualitative research, and has been evaluated for validity. During discussion among the review group, these tools were combined into a single instrument. These criteria were then applied to the all of the included studies after testing on a subgroup of 20 studies. The following criteria were amalgamated into a revised set of criteria.

For research papers:

- Is the statement of aim(s) of the paper clearly described?
- Is there congruity between the research methodology and the research questions or objectives?
- Is there congruity between the research methodology and the methods used to collect data?
- Is there a statement locating the researcher culturally or theoretically?
- Are participants and their voices adequately represented?
- Is the research ethical according to current criteria?
- Do the conclusions drawn in the paper appear to flow from the analysis or interpretation of data?

For non-research/textual papers (JBI 2011b):

- Is the statement of aim(s) of the paper clear?
- Is the source of the opinion clearly identified?
- Does the source of the opinion have standing in the field of expertise?
- Is the opinion's basis in logic/experience clearly argued?
- Is the argument that is developed analytical?
- Is there reference to the extant literature/evidence and any incongruence with it logically defended?
- Is the opinion supported by peers?

As with extraction, the principal benefit of a transparent and auditable trail of decision making with regard to how study quality was determined are key attributes of a high-quality systematic review (JBI 2011a). Within the timeframe requirements for this review, the need for transparency and auditability were given a high priority. The initial meetings to discuss the appraisal criteria, and the follow-up meeting after the test series had been completed were key aspects to protecting the quality and rigour of appraisal in this review.

2.3.4 Synthesis of evidence

Using the method employed by Oliver et al. (2008) in their framework synthesis as a guide, our review team used charts to create 'typologies' of the themes included in the revised conceptual framework by comparing two aspects of the framework at a time. This involved the creation of summaries of all the relevant literature for

each theme or concept included in the framework. These summaries, or 'typologies', were organised in charts, which were extracted from a larger synthesis table.

These expanded tables enabled our review team to map the impact of national and international assessment programmes on education policy and practice in developing countries by describing associations and relationships between and across themes in the framework. These charts further enhanced the transparency of the reported methodology used to derive conclusions (Dixon-Wood 2011).

Our synthesis framework, which was used to create typologies of themes (by exploring the relationship between aspects) and was then used to frame the description of relationships between themes, is provided in Appendix 2.4.

3. Identifying and describing literature

Outline of chapter

This chapter details the outcomes of searching for potential literature, screening and further excluding literature, and keywording and appraising the quality of included literature. The second part of the chapter provides a description of a subset of keyworded literature and examples of the included literature, as well as a description the quality and nature of the literature included for in-depth synthesis.

3.1 Identification, screening and keywording literature

Figure 3.1 systematically reports the outcomes of the identification of potential literature, screening literature, the application of the keywording and quality appraisal tools to the included literature, and the outcomes of the in-depth review and synthesis of the evidence.

3.1.1 *Identification of literature*

In total, 1,458 records were identified through our outlined search strategies (see sections 2.2.1 and 2.2.2) and were uploaded into the review software. After duplicates were removed, 1080 records with title, abstract and bibliographic citation were included in the review software for title and abstract screening.

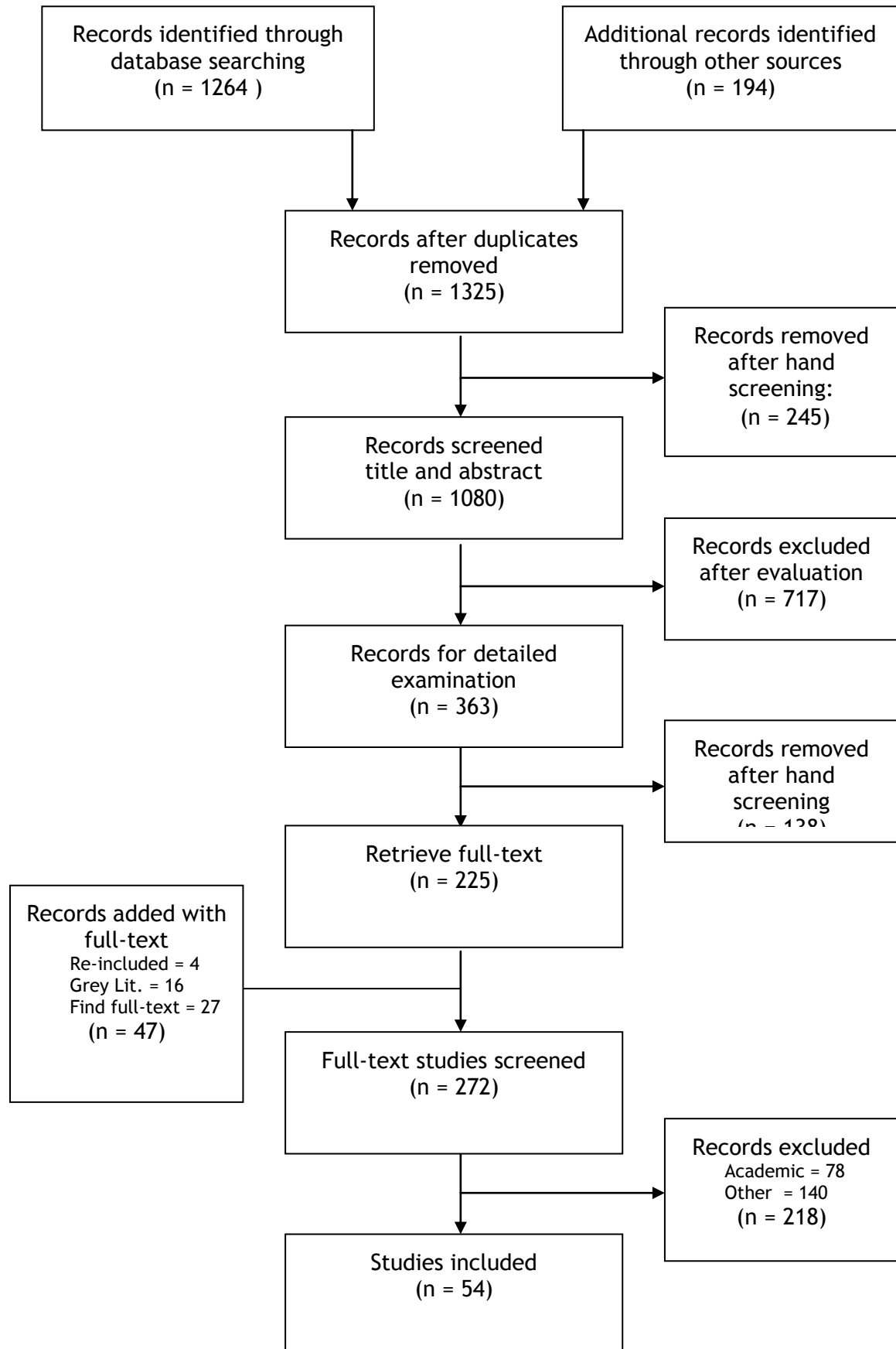
3.1.2 *Screening of titles and abstracts and full-text retrieval*

The 1080 records were screened on title and abstract according to the exclusion criteria (see section 2.2.3). From the title and abstract screening, 363 records were identified for full-text retrieval. After retrieving the full-text for the indicated records (see section 2.2.4), 272 records were included for descriptive mapping.

3.1.3 *Keywording of the literature*

The review team descriptively mapped the 272 records through the application of a keywording tool to the full-text of the included records. A further 140 records were excluded from the review based on relevance to the review question, using the exclusion criteria while 78 records were classified as academic papers (see also section 3.1.4) which left 54 records for in-depth review. The review team conservatively screened titles and abstracts during the initial screening process and included records when it was not possible to explicitly exclude a record based on the exclusion criteria. The availability of the full-text allowed the research team to further assess the relevance of the included literature, and exclude records which did not meet the inclusion criteria.

Figure 3.1: Systematic map: searching, screening, keywording and in-depth review



The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

3.1.4 Subset of included literature: papers with policy recommendations

As the nature and availability of the literature of interest to the review was not well known, the review team decided to include a subset of studies, largely academic papers, which analysed data from large-scale assessments and provided policy recommendations. The review team thought it of interest to include and describe the types of research that are thought to inform education policy-making, in order to introduce the main body of included literature. However, as these papers do not describe actual policy changes that have taken place, they are excluded from the main analyses for this study. A record included in this subset of literature analysed data from a standardised, large-scale assessment, reported factors that were associated with student achievement, and usually offered policy recommendations. Of the 272 records included for in-depth review, 78 records were coded as 'Academic papers with possible policy suggestions'. Appendix 3.1 reports in detail on the characteristics of the records included in this subset of the literature.

It is important to note that due to the nature of the included literature and the review software, there is not a one-to-one correspondence between the number of total records and coded keywords in this review. This is because one record may have reported on multiple countries, assessment programmes and policies. The review team decided that it was more important to keyword everything of relevance to the review that was included in a record, rather than prioritise a strict one-to-one correspondence between coded keywords and included records.

The subset of academic papers was keyworded to describe the nature of the assessment programmes (type of assessment programme, sampling approach, level of education assessed, region and country) and the types of policy recommendations made to policy-makers, who are one of the intended users of the research.

Examining the tables in Appendix 3.1, we can describe the types of assessment programmes and policy recommendations that were included in the academic papers relevant to our review. Considering the frequency of the regions that were keyworded across all included academic papers, the highest number of keyworded countries was in Africa -South of Sahara, Asia and then South America.

Describing the types of large-scale assessment programmes that were analysed to generate policy suggestions: most keyworded assessment programmes were national, and then regional in nature. The majority of large-scale assessment programmes used representative samples rather than census/population assessments, and assessed students in primary education (Years 1-6).

The subset of included academic papers also included policy suggestions based upon the analyses of large-scale assessment data. Overall, the majority of policy suggestions referred to resource allocation policies, particularly policies related to teacher in-service professional development, instructional materials and pre-service teacher preparation. Teaching and learning policies mainly highlighted in-class learning strategies and student oriented pedagogy. Other frequently coded policy suggestions related to curricular standards/reform, then assessment policy, parent engagement policy and performance standards.

3.1.5 Grey literature

From the 272 records which were descriptively mapped with the developed keywording code set, 140 were further excluded based on the exclusion criteria, 78 were coded as a subset of 'academic papers with possible policy suggestions' and 54 records were included for in-depth review and synthesis. Of the 54 included records, 27 were classified as 'grey literature'. This grey literature was located through professional networks and contacts, citation chasing and targeted website searches of and inquires to international agencies, research firms and national ministries of education. Bibliographic information for 50 records included for detailed synthesis, which also includes grey literature that is publicly available, are demarcated in the References chapter by an asterisk. The remaining four records included for detailed synthesis are grey literature that are not publicly available.

3.1.6 Examples of literature included for in-depth review

Crespo et al. (2000): This study analysed the impact of the Brazilian National System of Evaluation of Basic Education on education policies and practices, especially at the state level. It utilised data from interviews, site visits, document reviews, and a survey completed by state education departments. The study identified the strengths and weaknesses of the evaluation system.

Benveniste (2000): The author detailed the education system of Uruguay and the development of education reforms and the new national assessment programme through document analyses, surveys and interviews with education officials. The study outlined the associated facilitators for the government's construction of the assessment programme to garner public support, which resulted in successful implementation and use of the assessment data, despite a centralised education system.

Ferrer (2006): This study, supported by the Partnership for Educational Revitalization in the Americas (PREAL), provided summaries of 19 national and sub-national (state) education systems in North, Central and South America. The author conducted extensive policy document analyses and interviews with education officials from the 19 education assessment systems, and also supplemented primary data with published research. The study detailed the history of each assessment programme, assessment characteristics, educational policies and reform, and facilitators and barriers to the successful design, implementation and use of data from assessment programmes.

Gilmore (2005): This study examined the impact of participation in the Progress in International Reading Literacy Study (PIRLS) and TIMSS in low- and middle-income countries, serving as an evaluation of the value of World Bank support to these countries. Data were obtained by the author herself, and are drawn from the observation of four-day meetings of both PIRLS and TIMSS, interviews with the National Research Coordinator (NRC) of each World Bank-funded country, and responses to questionnaires sent to the NRC and a nominated senior education official of each country.

Leste (2005): This paper was a presentation by an official from the Ministry of Education of the Republic of Seychelles at a SACMEQ research conference. It described the way SACMEQ data were utilised at different stages of policy development (informing policy-makers, policy dialogue and policy action) resulting in policies against the streaming of students by abilities.

Mesa et al. (in press): This study examined the ways in which the use of international assessments, PISA and TIMSS impact on teaching and learning in the classroom through document analysis. For the countries relevant to this review,

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

Colombia and Indonesia, the study concluded that international assessments have not impacted on teaching and learning practices. PISA did influence the reform of curricular standards at the national level in Colombia, although decentralisation of curricular design and implementation to the local level in Colombia may lead to divergences in the interpretation and implementation of the reformed standards.

Nzomo and Makuwa (2006): This book chapter described the processes undertaken by the ministries of education in Kenya and Namibia in utilising SACMEQ findings to, among other things, undertake modifications to the curriculum, allocate funding to monitor education quality and develop programmes to bring greater efficiency to the education system. The authors are current and former NRCs in the two countries and extensively cite government policy documents.

A synthesis table of all studies that were included in this review for in-depth synthesis (excluding the four unpublished grey literature records) is included in Appendix 3.2. This appendix reports on coded assessment programme and policy impact characteristics by record.

3.1.7 Examples of literature excluded from the review

Braun and Kanjee (2006): This paper provided a framework of how national assessments can impact upon an education system and discussed the potential uses and impact of national assessments on education policy in developing countries. However, it did not make any reference to actual or existing links between national assessment findings and policy changes.

Lubisi and Murphy (2002): This article reviewed assessment policy and practice in South African schools in recent years and offered an overview of historical changes. The focus was on classroom-based assessments and not standardised assessments and was therefore excluded from this review.

Sunderman and Orfield (2008): This study examined the response of state education departments tasked with designing interventions in underperforming schools in the mandated yearly standardised tests in elementary and middle schools under the No Child Left Behind Act in the USA. The authors collected data from interviews by analysis of policy and programme documents, and from budget and staffing information. As it only referenced the USA, however, it was excluded from this review.

3.1.8 Description of literature in the review: review-specific data extraction

The review team applied a standardised data extraction tool to the included literature, in order to describe the literature that informed the review question.

Examining the frequency of applied keywords in Table 3.1, the included literature employs a wide breadth of methodologies to establish a link between the use of assessment data and education policy-making. As the review did not exclude literature based on design, it is expected that the included literature would use a variety of methods.

The review team analysed records which were coded as 'Other' and thematically grouped these records to form a new category not anticipated by the pre-defined tool: 'Data analysis and involvement in assessment programme'. This category described records which reported data analyses from the assessment programme, and linked this data to policy-making. These records often established the link to policy-making through their involvement in the assessment programme, for example, as the coordinating agency of the assessment programme.

The affiliation of the author, in relation to the production of the study, is also reported in Table 3.1. The included literature in the review was undertaken by a variety of bodies, notably international agencies and ministries of education - often principal stakeholders in large-scale assessments. It is also interesting to note that even though slightly over half of the included literature is 'grey' in nature, a considerable number of included records are from authors affiliated with universities.

Table 3.1 Description of included literature

Description of methodology	Frequency of applied keywords
Document analysis	23
Opinion/textual	14
Survey	14
Interviews/observations	13
Other :	
• Data analysis and involvement in assessment programme	
• Observation (e.g. attendance at meetings)	
Description of author's affiliation	
University	15
International agency	13
Ministry of education	12
Independent research organisation	11
Other government body	3

3.1.9 Quality assessment of literature in the review

The review team appraised the quality of the included literature using the adapted checklist from JBI, reported in Section 2.3.2. Using the checklist as a guide, the review team assigned a marker of quality to each record: either 'low quality' or 'high quality'. Approximately three-fifths of the included records were appraised as being of 'high quality' (n=31), while two-fifths of the included records were appraised as being of a 'low quality' (n=23).

To examine the specific characteristics of the literature that contributed to this appraisal, the review team reviewed the applied coding of the appraisal checklist. The review team assessed most uncertainty regarding the rigour of the literature in response to four criteria: the congruency between methodology and data collection, the extent to which records were situated in existing research, the conclusions drawn from analysis, and possible conflicts of interest. Where the team perceived there to be more potential of conflict of interest, there may be more uncertainty as to the soundness of the conclusions drawn in that literature. As many of the authors were primarily affiliated with ministries of education or international agencies, principal stakeholders in large-scale assessments, it follows that the team perceived more potential for possible bias.

4. In-depth review: results

Outline of chapter

This chapter reports the results of the in-depth review and provides a synthesis of the evidence. The first part of the chapter presents in-depth findings for the five sub-questions that were taken from the conceptual framework:

- i. What are the characteristics of the large-scale (i.e. national and international) assessment programmes that have informed education policy-making in developing countries?
- ii. How are the data from assessment programmes used to inform education policy-making in developing countries?
- iii. At what stages of the policy process are data from assessment programmes being used in developing countries?
- iv. What educational policies in developing countries have resulted from the use of assessment data?
- v. What are the facilitators and barriers to the use of assessment data in education policy-making?

The second part of the chapter reports the results of the synthesis of the themes identified in the conceptual framework, in order to address the main research question of the review:

What is the evidence of the impact of national and international assessment programmes on education policy-making, particularly policies regarding resource allocation and teaching and learning practices in developing countries?

4.1 Characteristics of assessment programmes used in education policy-making

To address the first sub-question of the review – i. *What are the characteristics of the large-scale (i.e. national and international) assessment programmes that have informed education policy-making in developing countries?* – this section reports on the characteristics of the assessment programmes that were used to inform education policy-making in the literature included in this review. Specifically, it describes the regions and countries, the types of assessment programmes, the sampling approaches and the levels of education assessed in the programmes. Tables for this section (Tables A4.1-A4.6) are in Appendix 4.1.

It must be borne in mind while interpreting these tables that there is not a one-to-one correspondence between the number of records and keywords across records, nor within records, as it was decided to keyword everything of relevance to the review.

Considering the geographical distribution of direct links between assessment programmes and education policy-making, Table A4.1 reports the frequency of countries and regions as coded across all of the included literature. The results demonstrate links between assessment programmes and education policy-making primarily in Africa - South of Sahara, South America and Asia.

Table A4.2 reports the frequency of types of assessment programmes as coded across all of the included literature. The majority of assessment programmes linked

to education policy-making in developing countries are primarily national in nature, then international and regional.

Examining the relationship between type of assessment programme and region (Table A4.3), national assessments overwhelmingly took place in South America, and to a lesser extent in Africa -South of Sahara and Asia, which reflects the overall geographical distribution of assessments in the included literature. Most of the regional assessments took place in Africa -South of Sahara, which reflects the importance of the regional assessments SACMEQ and PASEC (Le Programme d'Analyse des Systèmes éducatifs des États et gouvernements membres de la CONFEMEN [La Conférence des ministres de l'éducation des pays ayant le français en partage]) within Africa for education policy-making. The remaining regional assessment programmes primarily took place in North, Central and South America. This is probably due to LLECE's ongoing Regional Explanatory and Comparative Studies (i.e. Second [SERCE]/ Third [TERCE] Regional Comparative and Explanatory Study). Most international assessments were coded in conjunction with countries in South America and Asia, which may reflect the use of international assessments (principally PISA, TIMSS and PIRLS) in policy-making in these regions.

Table A4.4 reports the frequency of the assessment programme sampling approach as coded across all of the included literature. The assessment programmes primarily used a representative sampling approach. Rather than reflecting a relationship between the sampling approach of an assessment programme and goals of policy-making, i.e. as an accountability tool, the high frequency of keyworded representative sampling approaches is more likely indicative of the financial and technical constraints that face developing countries in implementing a large-scale assessment, as census/population assessments are very costly to undertake (see section 4.5.2 Barriers).

Table A4.5 shows that, in the included literature, most programmes assessed students in both primary and secondary education, then primary education only (up to Year 6), and least in secondary education only (Year 7 and above).

Table A4.6 indicates the relationship between level of education and type of assessment programme. National assessment programmes are overwhelmingly coded in conjunction with programmes that assess both primary and secondary education, then primary education only. Developing countries have expanded educational access across all levels of primary and secondary education, and even though student attrition rates typically increase in secondary education and educational access is still an issue, policy-makers are interested in educational quality and student outcomes at an aggregate systems level. Regional programmes are coded mostly in conjunction with primary education (the focus of SACMEQ, PASEC and LLECE studies). International programmes are coded primarily in conjunction with secondary education - largely due to 15-year-olds being the target population for PISA - and both primary and secondary education - largely due to students in Years 4 and 8 being the target population for TIMSS.

4.2 Goals and uses of assessment programmes

To address the second sub-question of the review – ii. *How are the data from assessment programmes used to inform education policy-making in developing countries?* – this section examines the goals and uses of the assessment programme data as evidence in policy-making. Appendix 4.2 reports the frequency of goals and uses of the assessment programmes as coded across all of the included literature. Almost all assessment programmes discussed in the literature reported multiple goals and uses. In one case, in Morocco, it was reported that the goal of the country's participation in international assessments changed from one of leverage

or as a mechanism of sanction, to one aimed at improving the quality of education (Gilmore 2005).

For the included literature, over one-third of assessment programmes were most often used to measure and ensure quality of the education system and its strengths and weaknesses by examining student achievement over time. Secondly, assessment programmes were coded, almost equally, as being used to measure and ensure equity and accountability. To measure and ensure equity, programmes were often intended to diagnose issues and target interventions in order to improve the educational outcomes for disadvantaged groups. To measure and ensure accountability, assessment programmes were often used to practise control over the system by reporting assessment results to relevant stakeholders internal or external to the education system. Though coded less frequently as a goal or use of the assessment programme, it is important to note that the *de novo* theme of leverage was coded as a primary goal in many studies. This goal, often in studies that critiqued the use of assessment programmes, indicated that the primary goal for the use of assessment programme data was to apply leverage to pre-existing political priorities.

Several other *de novo* themes emerged from a thematic analysis of the 'Other' keyword when the goals included in the keywording tool did not adequately describe the *primary* goal or intended use of the assessment data. These applied 'Other' codes can be grouped into four themes:

- To help inform future assessments and build technical capacity;
- To enable broad international comparisons;
- To provide inputs to be used at the local level for teachers, parents and students; and,
- To evaluate and examine the effects of pre-specified policies.

With regards to the fourth theme, there were several references in the literature to assessment programmes being used to monitor the progress of the educational system against specific goals or to evaluate the outcome of specific policies. Examples of the former include measuring progress towards the 'Education for All' goals (Chinapah 2000) and the Millennium Development Goals (Maligalig and Albert 2008). Regarding the evaluation of specific policies, this often occurred in assessment programmes, such as SACMEQ, which involve education policy-makers in the design of the assessment programme to ensure that relevant data for high priority issues are likely to lead to policy-making. The policies that are reported to be evaluated using assessment data included teacher training (Bernard and Michaelowa 2006, Greaney and Kellaghan 2008), performance-based incentives (Kellaghan et al. 2011, Mizala and Urquiola 2007) and ability grouping/streaming (Leste 2005).

4.3 Stages of the policy process

To address the third sub-question of the review – iii. *At what stages of the policy process are data from assessment programmes being used in developing countries?* – Appendix 4.3 reports the frequency of policy process stages as coded across all of the included literature. In the literature, the uses of assessment programme data were generally distributed across the stages of the policy process: in agenda setting, policy formulation, policy implementation and monitoring and evaluation. Overall, data from assessment programmes were used one-quarter of the time for monitoring and evaluation purposes. Even though the most frequent goal of the assessment programmes was to ensure quality (see section 4.2), which was anticipated to inform the agenda-setting and policy-formulation stages, assessment

data were used throughout the policy process, primarily for monitoring and evaluation purposes.

A fifth *de novo* stage of the policy process emerged in the preliminary keywording, ‘No impact on the policy process’. This code was applied only when there was an explicit mention of assessment data not impacting on policy-making; it was not applied to describe the absence of a discussion on policy-making in the literature. The code was applied to 13 records, which included all types of assessments in all regions (excluding the Middle East and the Pacific, which together accounted for only 3 percent of keyworded regions). Furthermore, authors who discussed this theme were associated primarily with international agencies, independent research organisations, then universities and ministries of education. Nine of the 13 records were assigned a marker of ‘high quality’ while the remaining four were assigned a marker of ‘low quality’ in terms of the technical quality of the study. Considering the varied characteristics of records that include this *de novo* theme, the review team did not perceive a potential for bias in the types of literature that discussed no impact. To better understand the factors that were associated with assessment data explicitly *not* having impacted on the policy process, the coded barriers for these programmes were collated and thematically grouped. This analysis resulted in four general themes that are discussed below:

- i. Assessment programme and analyses:* Barriers related to the soundness of the programme and analyses included poor-quality data, analyses unable to examine factors associated with student achievement to inform policy-making, no analyses of policy concerns, assessment not conducted regularly/long-term monitoring not coincident with the policy cycle, and policy concerns of developing countries not measured in international assessments.
- ii. Financial constraints:* The cost of conducting a large-scale assessment was often prohibitive and affected the ability to collect and analyse meaningful data for policy-makers; countries often did not have the sufficient financial resources to continue the assessment programme after funding from the external donor agency ended.
- iii. Weak assessment bodies and fragmented government agencies:* The capability of assessment agencies to conduct high-quality assessments was often affected when the agency was housed within the ministry of education. As a result, these agencies were not insulated from changing political agendas, bureaucracy, limited financial resources and high staff turnover. Furthermore, weak relationships between government agencies and departments, and the assessment agency and the ministry of education, often impeded the use of assessment results in informing policy-making.
- iv. Low technical capacity of assessment staff:* Low technical capacity of the national assessment staff often affected the ability to interpret and use the assessment results. Also, staff inexperience and low capacity for dissemination activities often impacted on their ability to influence policy-makers and promote the use of assessment data in policy-making.

Further barriers cited included political sensitivities to dissemination and resistance from teacher unions in the instance of poor performance.

4.4 Education policies resulting from the use of assessment data

To address the fourth sub-question of the review – iv. *What education policies in developing countries have resulted from the use of assessment data?* – Appendix 4.4 reports the frequency of education policies as coded across all of the included literature. The policies are reported first under Table A4.7 ‘Resource allocation’,

then Table A4.8 ‘Teaching and learning practices’ and finally all other systems-level policies under Table A4.9.

The most frequently occurring resource allocation policies that resulted from the use of assessment data were those intended to improve the quality of the teacher and teaching materials used in schools, which were seen as educational inputs into the system, in addition to those that made changes in education funding. ‘In-service professional development’ policies were intended to improve the quality of the teacher by offering improved, targeted, increased frequency, or new delivery professional development programmes. For example, new in-service professional development programmes were offered online for teachers and leaders in Brazil (Martinez 2007); and in-service programmes for teachers in Macedonia emphasised new pedagogical practice (student-oriented pedagogy) (Elley 2005).

Teacher preparation policies aimed to improve and increase teacher qualifications, training and experience before they entered the education system. Policies for ‘Instructional materials’ focused on the materials used by teachers and students in the classroom. For example, this type of policy most often referred to textbook provision or textbook reform, as in Kyrgyzstan (Shamatov and Sainazarov 2010).

Changes to education funding were intended to improve educational outcomes by providing: funds for compensatory school interventions and programmes for low-performing and low-SES (socio-economic status) schools, as in Chile (Ferrer 2006); performance-based financial incentives for schools and teachers, as in India (Mukhopadhyay and Sriprakash 2010); and changed funding allocation between schooling sectors (public and private).

A thematic analysis of the ‘Other’ *de novo* resource allocation policies that were not included in the coding tool is presented below. These resource allocation policies were quite diverse, and included:

- Financial incentives for private industry investment in public education/private sector partnerships (for STEM [science, technology, engineering and mathematics] industry and high-performing schools);
- School health programmes;
- School transportation programmes;
- Resource standards and benchmarks for inputs; and,
- Introduction of multi-grade classrooms.

Examining Table A4.7, the most frequently occurring policies related to teaching and learning practices were ‘Student-oriented pedagogy’ and ‘In-class learning strategies’. These policies specifically focused on affecting student learning in the classroom through such strategies as differentiated learning, collaborative learning, or increased practical work such as seen in experiments in science teaching in Malaysia (Gilmore 2005).

Even though the ‘Other’ code was the most frequently applied in Table A4.8, this code covered a wide array of policies that were not contained in the keywording tool. A thematic analysis of the ‘Other’ *de novo* teaching and learning policies revealed that, primarily, assessment data were used to develop and disseminate teacher and leadership guides (information, guidelines and frameworks) for: curricular topics where teachers may have misconceptions; recommended pedagogic practices to target knowledge/skills assessed by ‘difficult items’; curricular lesson planning; selection of classroom texts; checklists for identified good practice/management; and implementation of practical classroom investigations.

The remaining ‘Other’ teaching and learning policies included the development and dissemination of classroom assessment frameworks for teachers.

Therefore, within ‘Teaching and learning practices’, policies related to student-oriented pedagogy and learning strategies were the most frequent policies resulting from the use of assessment data as evidence in policy-making, frequently disseminated through teacher and leadership guides

Other education policies that were informed by the background literature, or were *de novo* themes that emerged from the preliminary keywording exercise, were also included and coded. These system-level policies are presented in Table A4.9 in Appendix 4.4. The policies that most frequently resulted from the use of assessment data were related to curriculum standards and reform, followed by those related to assessment programmes and performance standards. Even though, thematically, resource allocation policies were the most frequently occurring in policy-making, ‘Curriculum standards/reform’ was the most frequently occurring policy overall.

A *de novo* barrier to the effective use of assessment data in policy-making was noted in the literature: ill-defined curricula and curricular standards impeded the development of a meaningful assessment framework and analyses as well as the interpretation of the results. Therefore, not only would poor academic results influence the development of curricular reform and standards, but the implementation of an assessment programme itself necessitated a clearly defined curricular framework within which to operate and communicate with relevant stakeholders. It is therefore not surprising to also see that policies relating to assessment programmes and performance standards were the most frequently occurring policies after curriculum reform and standards.

The use of assessment data in policy-making often informed the creation of an assessment policy. In the literature, these assessment policies often established national assessment bodies or agencies and legislation to legally mandate assessments, and created action plans for future systematic implementation of assessment programmes and policies related to the use of assessment data.

Performance standards would also help policy-makers, teachers, parents and other stakeholders to interpret and *use* the assessment results in a meaningful way. A thematic analysis of the ‘Other’ *de novo* system-level policies that were not included in the coding tool is presented below. The applied ‘Other’ codes for system-level policies could be grouped into two themes:

- Reduction or prohibition of the use of grade-repetition policy and use of automatic promotion; and,
- The use of assessment data to target low or high performing schools, or disadvantaged groups for targeted whole-school interventions. These multi-faceted interventions targeted various factors theorised to affect student achievement: professional development, resources, pedagogy, leadership and school finance (Please also see discussion of education funding at the beginning of section 4.4).

4.5 Facilitators and barriers to the use of assessment data

To address the fifth sub-question of the review – v. *What are the facilitators and barriers to the use of assessment data in education policy-making?* – Appendix 4.5 reports the frequencies of the facilitators and barriers as coded across all of the included literature.

4.5.1 Facilitators

Examining Table A4.10, the most frequently cited facilitators to the use of assessment data were media and public opinion and the dissemination of appropriate results to stakeholders, followed by the soundness of the assessment

programme itself and how well the programme was integrated into policy processes.

Media and public support were principal factors associated with the use of large-scale assessment data in policy-making. Extensive media coverage of assessment programme results and public opinion, often in the wake of poor learning outcomes, can create a 'shock window', as in the case of PISA results (Breakspear 2012). For example, Chile's public dissemination and media coverage of national assessment programme results has served as a primary mechanism to enact reforms as education policy-makers and managers are externally accountable to the public and the government (Meckes and Carrasco 2010).

The dissemination of appropriate results to stakeholders was also a primary facilitator to the use of assessment data in policy-making. From the literature, appropriate results to stakeholders encompassed two broad aspects: dissemination to a wide variety of stakeholders at all levels of possible education reform, as well as differentiated dissemination appropriate to the stakeholder group. Considering the variety of stakeholders, Arancibia (1997) notes that from the experience of LLECE studies, dissemination to a wide variety of stakeholders at national, sub-national and local levels will increase the likelihood that assessment data will inform policy.

The second aspect, differentiated dissemination, focuses on appropriate targeting of results by stakeholder group (Greaney and Kellaghan 2008). For example, effective models of assessment reporting frameworks to students and parents can present information in a variety of ways to make data more relevant and useful (Ferrer 2006). Successful models for teacher reporting, from Uruguay, have included selection of test items and scoring guidelines for teachers to create their own class-based assessment tools and frameworks. Dissemination to national and sub-national policy-makers in Senegal was conducted in face-to-face seminars to start a policy dialogue (Bernard and Michaelowa 2006).

The soundness of the assessment programme encompassed various aspects. From the literature these included: trustworthy and reliable data; criterion-referenced assessment programmes; clearly defined frameworks and standards within the assessment; and regular implementation of assessment over time.

Integration of the assessment programme into policy processes also encompassed various aspects. To illustrate with examples from the literature: the PASEC international assessment prioritised regional policy concerns in the design of the assessment which made it more likely that results would be used in policy-making, as in the case of Malawi (Saito and vanCapelle 2010); legislated assessment programmes provided a mandate for programmes to be regularly conducted, well-designed and used in policy-making; an established and well-respected assessment agency *within* the Ministry of Education in Chile (Ferrer 2006) helped the assessment body to remain insulated from political instability while simultaneously allowing the assessment to directly respond to policy concerns and priorities of the government. However, it must be highlighted that *weak* assessment agencies within ministries of education were associated factors in cases of no impact on policy-making (see section 4.3). Similar to Chile, feedback loops between the assessment agency and government in Jordan helped to facilitate impact on policy-making (Abdul-Hamid et al. 2011).

Two new important themes arose from the analysis of facilitating factors initially placed in the 'Other' category. These related to funding and continuity/stability. With regards to funding, the literature described budgetary increases for the Brazilian assessment programme (Castro 2010), the adequacy of funding to conduct

national assessment in Uganda (Kanjee and Acana 2010), the strategy in Togo to associate the goals and activities related to participation in regional assessment with international donors' goals and activities to ensure continuing financial support (Bernard and Michaelowa 2006) as examples of facilitating factors. The literature described how the continuity and stability of regular cycles for conducting assessments in Brazil (Castro 2010), Chile, Colombia (Ferrer 2006) and Uganda (Kanjee and Acana 2010, Ravela 2005) facilitated their use in policy-making.

Other facilitating factors included:

- Cross-country comparisons from international or regional assessments; and,
- Publicly available databases for further research and dialogue (PASEC - Bernard and Michaelowa 2006).

4.5.2 Barriers

Table A4.11 reports the most frequently coded barriers to the use of assessment programme data in policy-making. The most frequently coded barriers were related to the quality of the assessment programme itself, i.e. soundness of the assessment programme and, furthermore, in-depth analysis of the assessment data and diagnosis of issues. The following are specific examples from the literature to contextualise issues of assessment programme quality and analyses which acted as barriers to the use of data in policy-making:

- Identified problems with test items led to misinterpretation of results by policy-makers;
- Assessed skills and content were not comparable over time which led to misinterpretation of educational quality from one assessment cycle to the next;
- Assessments were not well designed for facilitating meaningful comparisons (to measure and ensure equity);
- Assessments were not responsive to the pressing policy concerns of the education system;
- Implementation of the assessment programme was infrequent;
- Assessment data were unable to provide information to policy-makers regarding extent (e.g. time, amount) of inputs needed for reform;
- In-depth analyses were limited due to financial constraints; and,
- Assessment results were de-legitimised due to the low quality of assessment sampling and administration.

After issues related to assessment programme quality, the most frequently coded barriers to the use of assessment programme data in policy-making were linked to dissemination activities, specifically dissemination to the public, as well as other stakeholders. In most of these cases, dissemination to the public and other stakeholders was inadequate, as for example with the national assessment programme in the Philippines (Maligalig and Albert 2008), or data were not relevant to the needs of stakeholders. Furthermore, inadequate dissemination to the public, and erroneous public expectations for the assessment programme and resulting reforms, led to the inappropriate use of data in policy-making (Gutiérrez and Vázquez 2008) or discouraged effective long-term policies in favour of quick results (Meckes and Carrasco 2010).

4.6 Synthesis of the evidence

Results presented in the previous section were based on an analysis of one aspect at a time. Thus, data were compiled and discussed separately for each of the following topics:

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

- i. Types of assessment programmes – that is the characteristics of the large-scaled assessment programs as per first of this review's sub-questions;
- ii. Goals of the assessment – that is how data from large-scaled assessment programs are used as per the second of this review's sub-questions;
- iii. Stage of the policy process – that is at what stages of the policy process data from large-scaled assessment programs are used as per the third of this review's sub-questions;
- iv. Types of policies – that is the resulting educational policies in developing countries as per the fourth of this review's sub-questions; and,
- v. Facilitators and barriers – that is the facilitators and barriers to the use of assessment data in education policy-making as per the fifth of this review's sub questions.

The aim of this exercise was to allow us to address the main research question:

What is the evidence of the impact of large-scale assessment programmes on education policy-making, particularly policies regarding resource allocation and teaching and learning practices in developing countries?

This section goes one step further and synthesises the results.

In the synthesis, data from different aspects are examined together by cross-tabulation in order to identify potential patterns in the data. The cross-tabulations are guided by the synthesis table (see Appendix 2.4). However, given the amount of data available, an analysis of the complete table with all its aspects would not provide meaningful results, as the number of data points in each of the cells would be too small. In addition, it is impossible to uniquely locate data points in this table as documents in the review frequently talk about, for example, the fact that the topic of equity is used both in agenda setting and for monitoring and evaluation.

Therefore, the synthesis was undertaken by cross-tabulating two aspects at a time. Given the overarching research question with its focus on policies and resource allocation and teaching and learning practices, this was done by cross-tabulating first the policy goals (i.e. quality, equity, accountability and leverage) and stages of the policy process (i.e. agenda setting, policy formulation, policy implementation and monitoring and evaluation) and second the types of policies (i.e. research allocation and teaching and learning practices) with each of the following: (i) assessment type (i.e. sub-national, national, regional, international), (ii) region; (iii) education level (i.e. primary, secondary, both); and (iv) sampling approach (i.e. sample, census).

4.6.1 Policy goals - synthesis

Results of the first step of the synthesis, cross-tabulating the policy goals of quality, equity, accountability and leverage with assessment type, region, education level and sampling approach, are given in Tables 4.1 to 4.4.

Table 4.1: Policy goals and assessment type

Policy goal	Sub-national (state-level) (n=4)*	National (n=32)*	Regional (n=16)*	International (n=17)*
Quality	4	25	15	14
Equity	3	16	11	8
Accountability	4	19	8	7
Leverage	1	8	4	6
Other	0	7	2	3

*n = total count of applied codes in the in-depth analyses of the sub-questions.

Count in cells indicate frequency of cross-coding.

Cases of 'No impact' and 'Unknown' are not included.

Table 4.1 illustrates that national assessments were mentioned more frequently in connection with quality as a policy goal than regional and international assessments. Similarly, national assessment programmes tended to inform the policy goals of equity and accountability more than regional and international assessment programmes. However, these results are probably influenced by the fact that more documents (n=31) refer to national assessment programmes than regional (n=16) and international (n=21) programmes. In this context, it is interesting to note that, while small in absolute numbers, the sub-national assessments (n=4) feature relatively far more prominently than the other assessment types when it comes to the policy goals of quality, equity and accountability.

Table 4.2 Policy goals and region

Policy goal	Europe (n=5)*	Africa - North of Sahara (n=4)*	Africa - South of Sahara (n=66)*	North and Central America (n=18)*	South America (n=49)*	Asia (n=24)*	Middle East (n=2)*	Pacific (n=3)*
Quality	4	2	17	6	19	13	2	1
Equity	2	1	11	3	13	8	1	1
Accountability	1	0	11	2	14	7	1	0
Leverage	2	1	7	0	8	5	2	0
Other	2	1	5	1	8	4	1	1

*n = total count of applied codes in the in-depth analyses of the sub-questions.

Count in cells indicate frequency of cross-coding.

Cases of 'No impact' and 'Unknown' are not included.

The results of the cross-tabulation between policy goals and region are given in Table 4.2. These show that, in absolute terms, quality as a policy goal is mentioned almost equally in Africa -South of Sahara (n=17) and South America (n=19). However, in relative terms, the occurrence is far greater for South America (total n=49 and also for Asia [n=13 of a total of n=24] than for Africa -South of Sahara (total n=75) when the total number of applied codes is taken into consideration.

Furthermore, it is interesting to note that, for each region, quality is the policy goal most frequently associated with large-scale assessments, whereas equity, accountability and leverage receive less mention across all regions.

Lastly, the relatively high occurrences of ‘Other’ goals of assessment programmes in South America appear to warrant further examination. A thematic analyses of the applied ‘Other’ code revealed three broad goals not adequately captured in the keywording tool. The first goal was to specifically use international and regional assessment programmes to comparatively assess the quality of education in a broad, international context. The second goal of assessment programmes noted, from a critical perspective, that assessment programmes were implemented as a result of direct ‘policy borrowing’ from developed countries without regard to the needs or goals of the education system. The third goal was for assessment programmes to provide useful information and inputs directly to teachers, parents and students. These inputs were not intended to promote accountability but rather to foster a local culture of evaluation in schools.

Table 4.3: Policy goals and education level

Policy goal	Primary education (n=25)*	Secondary education (n=15)*	Both primary and secondary education (n=27)*
Quality	21	11	21
Equity	12	5	13
Accountability	10	5	16
Leverage	7	4	9
Other	5	2	6

*n = total count of applied codes in the in-depth analyses of the sub-questions. Count in cells indicate frequency of cross-coding.

Table 4.3 indicates that quality is again the policy goal that features the most, in this case in relation to education levels. In addition, it is apparent that the relative importance of policy goals do not differ depending on the education level. Thus, for example, quality is identified as a policy goal most frequently, in primary, secondary and both primary and secondary education levels. The only exception to this finding is that accountability as goal features relatively more often where assessments occur at both the primary and secondary school level. Table 4.3 also shows that all policy goals are mentioned far more frequently together when both primary and secondary education are the target for assessment than for the two education levels being assessed separately.

Table 4.4: Policy goals and sampling approach

Policy goal	Census (n=17)*	Representative sample (n=43)*
Quality	13	35
Equity	11	21
Accountability	12	18
Leverage	5	12
Other	5	9

*n = total count of applied codes in the in-depth analyses of the sub-questions. Count in cells indicate frequency of cross-coding.

Table 4.4 shows that assessments that use a census approach and obtain information from each student in the population mention the policy goals of quality, equity and accountability almost equally. For assessments that collect data from a sample that is representative of the population, policy goals of equity and accountability are mentioned less often than quality.

4.6.2 Stages of the policy process -synthesis

As the next step in the synthesis, the stages of the policy process, i.e. agenda setting, policy formulation, policy implementation and monitoring and evaluation were cross-tabulated with assessment type (Table 4.5), region (Table 4.6), education level (Table 4.7), and sampling approach (Table 4.8).

Table 4.5: Policy process and assessment type

Stage of policy process	Sub-national (state-level) (n=4)*	National (n=32)*	Regional (n=16)*	International (n=17)*
Agenda setting	2	16	12	9
Policy formulation	1	11	7	10
Policy implementation	3	16	7	13
Monitoring and evaluation	2	20	11	13
No impact on policy process	2	7	3	7
Other	2	3	0	3

*n = total count of applied codes in the in-depth analyses of the sub-questions. Count in cells indicate frequency of cross-coding.

Table 4.5 reveals that across assessment types, large-scale assessments least impact upon policy formulation. Thus, regardless of whether an assessment is sub-national, national, regional or international, data are used slightly more in policy agenda setting, policy implementation and monitoring and evaluation than in the creation of policies. This means that large-scale assessments have the least impact on the ways in which analytical and political options and strategies for education policies are constructed (i.e. policy formulation). This finding also holds for the different assessment types, if we take into account the more frequent occurrence of national assessments compared with international, regional and sub-national assessments. In other words, just because sub-national and national assessment programmes are more accessible for a country's policy-makers, this does not mean that data from them are more likely to impact upon the probably more localised mechanisms involved in policy formulation.

International assessments, comparatively, have less impact on agenda setting and creating awareness of issues than national or regional assessment programmes. This may be because many countries that participate in international assessments already have an understanding of education issues from their own national assessments. On the other hand, international assessments impact more than national or regional assessments on policy implementation, and are used to inform how policies are targeted and implemented.

Table 4.6: Policy process and region

Stage of policy process	Europe (n=5)*	Africa - North of Sahara (n=4)*	Africa - South of Sahara (n=66)*	North and Central America (n=18)*	South America (n=49)*	Asia (n=24)*	Middle East (n=2)*	Pacific (n=3)*
Agenda setting	2	0	10	4	13	6	1	1
Policy formulation	2	1	10	3	13	7	1	0
Policy implementation	3	1	10	3	15	11	1	0
Monitoring and evaluation	3	2	11	3	16	11	1	0
No impact on policy process	2	1	5	4	9	6	0	0
Other	0	0	1	0	4	3	0	0

*n = total count of applied codes in the in-depth analyses of the sub-questions. Count in cells indicate frequency of cross-coding.

Table 4.6 indicates that there are no noteworthy differences in the impact of large-scale assessments on policy stages in different regions. Each region seems to mirror the general trend for the overall frequency which shows that assessment data impact equally on all stages of the policy cycle. A possible exception is Asia, for which a somewhat lower impact on agenda setting and policy formulation can be observed.

Table 4.7: Policy process and education level

Stage of policy process	Primary education (n=25)*	Secondary education (n=15)*	Both primary and secondary education (n=27)*
Agenda setting	14	9	14
Policy formulation	10	10	10
Policy implementation	13	8	15
Monitoring and evaluation	15	12	15
No impact on policy process	5	5	6
Other	3	3	2

*n = total count of applied codes in the in-depth analyses of the sub-questions. Count in cells indicate frequency of cross-coding.

Table 4.7 provides the results of the cross-tabulation of the stages of the policy cycle with education level. The impact on the different stages of the policy process

is similar for large-scale assessments at the primary and secondary education levels and the primary and secondary education levels combined.

Table 4.8: Policy process and sampling approach

Stage of policy process	Census (n=17)*	Representative sample (n=43)*
Agenda setting	10	22
Policy formulation	4	20
Policy implementation	10	22
Monitoring and evaluation	11	26
No impact on policy process	3	11
Other	2	4

*n = total count of applied codes in the in-depth analyses of the sub-questions. Count in cells indicate frequency of cross-coding.

The impact on the stages of the policy process is very similar for large-scale assessments, regardless of whether they use a census or a representative sample approach. Thus, as can be seen in Table 4.8, they are most used for agenda setting and policy implementation as well as monitoring and evaluation and, to a lesser extent, policy formulation. However, given the relative frequency of both approaches, assessments using a representative sample approach impact relatively more than assessments using a census sample approach on the policy formulation stage of the policy process.

4.7 Types of policy-synthesis

The main research question asked about not only the impact of large-scale assessments on education policy in general but on policies regarding resource allocation and teaching and learning practices more specifically. Hence, the synthesis also focused specifically on these two types of policies.

In this section, cross-tabulations are provided for assessment type, region, education level and sampling approach, first with resource allocation (Tables 4.9 to 4.12) and then with teaching and learning practices (Tables 4.13 to 4.16), and lastly with other education policy types (Tables 4.17 to 4.20).

In general, the large number of sub-categories of policies for resource allocation leads to relatively low numbers in each cell, which, in turn, makes it more difficult to identify potential patterns.

4.7.1 Resource allocation - synthesis

Table 4.9: Resource allocation and assessment type

Resource allocation	Sub-national (state-level) (n=4)*	National (n=31)*	Regional (n=16)*	International (n=21)*
Instructional materials	1	10	5	7
School supplies	0	2	2	1
Equipment	0	3	2	1
Facilities	2	4	5	2
School feeding/meals	1	2	1	2
Class size/ratios	0	0	2	1
Instructional time/school hours	1	1	1	1
Teacher preparation	1	6	4	3
Teacher recruitment and retention	2	4	4	2
In-service professional development	3	11	5	6
Funding formula	3	9	4	3
Decision-making authority	1	3	0	2
Other resource allocation policy	2	6	4	3

*n = total count of applied codes in the in-depth analyses of the sub-questions.
Count in cells indicate frequency of cross-coding.

The cross-tabulation in Table 4.9 between policies for resource allocation and assessment type largely reflects the overall descriptive results reported for resource allocation whereby the policies cited most often concern teacher quality (i.e. professional development, as well as preparation, recruitment and retention), instructional materials and the funding formula. This overall pattern seems to be repeated for the resource allocation sub-categories in sub-national, national, regional and international assessment types with no noteworthy differences emerging.

Table 4.10: Resource allocation and region

Resource allocation	Europe (n=5)*	Africa - North of Sahara (n=4)*	Africa - South of Sahara (n=66)*	North and Central America (n=18)*	South America (n=49)*	Asia (n=24)*	Middle East (n=2)*	Pacific (n=3)*
Instructional materials	1	0	7	2	7	5	1	0
School supplies	0	0	2	0	1	1	0	0
Equipment	0	0	2	0	1	2	0	0
Facilities	0	0	5	1	3	2	0	0
School feeding/meals	0	0	0	1	2	0	0	0
Class size/ratios	0	0	2	0	0	0	0	0
Instructional time/school hours	0	0	1	0	1	2	0	0
Teacher preparation	1	0	6	0	4	3	1	0
Teacher recruitment and retention	0	0	3	2	4	2	0	0
In-service professional development	0	0	7	1	8	7	1	0
Funding formula	0	0	4	2	10	3	0	0
Decision-making authority	0	0	2	0	3	3	0	0
Other resource allocation policy	0	0	4	2	7	2	0	0

*n = total count of applied codes in the in-depth analyses of the sub-questions.
Count in cells indicate frequency of cross-coding.

Table 4.10 provides results of the cross-tabulation between resource allocation and regions. There is some evidence for a greater impact on resource allocation policies related to funding formulae in South America than in other regions. In addition, somewhat greater impact on resource allocation policies regarding teacher preparation in Africa -South of Sahara emerges when compared with South America and Asia.

Table 4.11: Resource allocation and education level

Resource allocation	Primary education (n=25)*	Secondary education (n=15)*	Both primary and secondary education (n=27)*
Instructional materials	11	7	8
School supplies	2	1	1
Equipment	3	1	1
Facilities	4	1	4
School feeding/meals	0	1	2
Class size/ratios	1	0	1
Instructional time/school hours	1	1	2
Teacher preparation	5	3	6
Teacher recruitment and retention	3	2	5
In-service professional development	10	6	8
Funding formula	6	3	8
Decision-making authority	2	2	2
Other resource allocation policy	4	3	7

*n = total count of applied codes in the in-depth analyses of the sub-questions. Count in cells indicate frequency of cross-coding.

Table 4.11 provides the results of the cross-tabulation between level of education and policies for resource allocation. No noteworthy differences in the impact of assessment at different education levels on resource allocation are discernible, except for the slightly higher occurrence of in-service professional development and instructional materials for assessments at the primary education level compared with the secondary education level and the combined primary and secondary education levels.

Table 4.12: Resource allocation and sampling approach

Resource allocation	Census (n=17)*	Representative sample (n=43)*
Instructional materials	5	16
School supplies	0	3
Equipment	1	4
Facilities	3	6
School feeding/meals	1	1
Class size/ratios	0	2
Instructional time/school hours	1	2
Teacher preparation	3	9
Teacher recruitment and retention	3	6
In-service professional development	7	13
Funding formula	6	10
Decision-making authority	2	3
Other resource allocation policy	5	7

*n = total count of applied codes in the in-depth analyses of the sub-questions.
Count in cells indicate frequency of cross-coding.

Results reported in Table 4.12 for the cross-tabulation between resource allocation and sampling approach generally support the overall frequencies for resource allocation. However, in-service professional development and funding formula are mentioned more often in census assessments than in sample assessments when the much more frequent occurrence of the latter is taken into account.

4.7.2 Teaching and learning practices - synthesis

Table 4.13: Teaching and learning practices and assessment type

Teaching and learning practices	Sub-national (state-level) (n=4)*	National (n=31)*	Regional (n=16)*	International (n=21)*
Classroom management and discipline	0	0	0	0
In-class learning strategies	1	3	1	4
Student-oriented pedagogy	1	2	1	6
Enhanced learning activities	1	1	1	2
Staff collaboration/mentoring	1	2	0	0
Student-teacher relationships	0	0	0	0
Teacher attitudes	0	0	0	0
Organisation of instructional/study time	1	1	0	1
Additional classes	1	1	1	0
Extra-curricular activities	0	0	0	0
Motivation and future plans	0	2	0	2
Tracking/streaming policy	0	0	1	1
Other teaching and learning policy	2	7	5	5

*n = total count of applied codes in the in-depth analyses of the sub-questions. Count in cells indicate frequency of cross-coding.

Results of the cross-tabulation between teaching and learning practices and assessment type shown in Table 4.13 do not reveal any particular pattern. However, the most frequently reported types of teaching and learning practices involve in-class learning strategies and student oriented pedagogy, as well as policy practices coded as 'Other'. Upon further analysis, 'Other' was found to include mainly policies for the development and dissemination of teacher and leadership guides, and classroom assessment frameworks.

Table 4.14: Teaching and learning practices and region

Teaching and learning practices	Europe (n=5)*	Africa - North of Sahara (n=4)*	Africa - South of Sahara (n=66)*	North and Central America (n=18)*	South America (n=49)*	Asia (n=24)*	Middle East (n=2)*	Pacific (n=3)*
Classroom management and discipline	0	0	0	0	0	0	0	0
In-class learning strategies	0	0	1	3	3	2	0	0
Student-oriented pedagogy	2	1	2	2	3	4	0	0
Enhanced learning activities	1	0	0	1	1	1	0	0
Staff collaboration/mentoring	0	0	1	0	2	1	0	0
Student-teacher Relationships	0	0	0	0	0	0	0	0
Teacher attitudes	0	0	0	0	0	0	0	0
Organisation of instructional/study-time	1	1	2	0	2	2	0	0
Additional classes	0	0	1	0	1	1	0	0
Extra-curricular activities	0	0	0	0	0	0	0	0
Motivation and future plans	0	0	1	0	2	1	0	0
Tracking/streaming policy	1	0	1	1	1	1	0	0
Other teaching and learning policy	2	1	5	2	7	3	1	0

*n = total count of applied codes in the in-depth analyses of the sub-questions.
Count in cells indicate frequency of cross-coding.

Results of the cross-tabulation between teaching and learning practices and region in Table 4.14 shows no noteworthy regional impact of assessment on teaching and learning practices, except for a slightly greater impact on teaching and learning policies which emphasise student-oriented pedagogy in Asia.

Table 4.15: Teaching and learning practices and education level

Teaching and learning practices	Primary education (n=25)*	Secondary education (n=15)*	Both primary and secondary education (n=27)*
Classroom management and discipline	0	0	0
In-class learning strategies	1	2	3
Student-oriented pedagogy	2	3	3
Enhanced learning activities	0	1	2
Staff collaboration/mentoring	2	0	2
Student-teacher relationships	0	0	0
Teacher attitudes	0	0	0
Organisation of instructional/study time	2	0	2
Additional classes	2	0	1
Extra-curricular activities	0	0	0
Motivation and future plans	1	2	2
Tracking/streaming policy	1	1	0
Other teaching and learning policy	7	3	7

*n = total count of applied codes in the in-depth analyses of the sub-questions.
Count in cells indicates frequency of cross-coding.

The results of the cross-tabulation between teaching and learning practices and education levels, reported in Table 4.15, do not reveal any differences in the impact on teaching and learning practices from the education level at which assessments were applied.

Table 4.16: Teaching and learning practices and sampling approach

Teaching and learning practices	Census (n=17)*	Representative sample (n=43)*
Classroom management and discipline	0	0
In-class learning strategies	3	4
Student-oriented pedagogy	2	5
Enhanced learning activities	1	1
Staff collaboration/mentoring	2	2
Student-teacher relationships	0	0
Teacher attitudes	0	0
Organisation of instructional/study time	1	2
Additional classes	1	1
Extra-curricular activities	0	0
Motivation and future plans	1	2
Tracking/streaming policy	0	2
Other teaching and learning policy	6	10

*n = total count of applied codes in the in-depth analyses of the sub-questions. Count in cells indicate frequency of cross-coding.

The results reported in Table 4.16, which show results of cross-tabulation between teaching and learning practice and the type of sampling approach taken in assessments, again reveal no noteworthy deviations from the overall frequencies.

4.7.3 System-level policy types - synthesis

Table 4.17: Other policy type and assessment type

Policy type	Sub-national (state-level) (n=4)*	National (n=31)*	Regional (n=16)*	International (n=21)*
Assessment policy	2	11	7	11
Curriculum standards	2	12	6	14
Performance standards	1	9	9	9
Community/parent engagement policy	0	5	2	4
Accountability policy	0	7	5	2
Other	0	12	7	7

*n = total count of applied codes in the in-depth analyses of the sub-questions. Count in cells indicate frequency of cross-coding.

As well as policies for resource allocation decisions and teaching and learning practices, the study also identified a number of categories for other system-level education policies that may be impacted by large-scale assessments. The cross-tabulation between these other policies and assessment type, summarised in Table 4.17, indicated that the most frequently cited policy impacts are on curriculum standards, performance standards and assessment policy. Notably, two-thirds of all included studies involving international assessments reported an impact on curriculum standards.

Table 4.18: Other policy type and region

Policy type	Europe (n=5)*	Africa - North of Sahara (n=4)*	Africa - South of Sahara (n=66)*	North and Central America (n=18)*	South America (n=49)*	Asia (n=24)*	Middle East (n=2)*	Pacific (n=3)*
Assessment policy	4	1	6	2	10	7	2	1
Curriculum standards	3	1	7	3	13	10	1	0
Performance standards	2	0	8	2	8	5	1	0
Community/parent engagement policy	0	0	2	2	5	0	0	0
Accountability policy	0	0	7	0	4	4	0	0
Other	4	1	11	4	14	7	1	0

*n = total count of applied codes in the in-depth analyses of the sub-questions.

Table 4.19: Other policy type and education level

Policy type	Primary education (n=25)*	Secondary education (n=15)*	Both primary and secondary education (n=27)*
Assessment policy	11	6	10
Curriculum standards	10	13	10
Performance standards	8	8	8
Community/parent engagement policy	2	4	4
Accountability policy	7	2	5
Other	9	6	13

*n = total count of applied codes in the in-depth analyses of the sub-questions. Count in cells indicate frequency of cross-coding.

Table 4.20: Other policy type and sampling approach

Policy type	Census (n=17)*	Representative sample (n=43)*
Assessment policy	8	16
Curriculum standards	5	20
Performance standards	6	16
Community/parent engagement policy	2	8
Accountability policy	5	8
Other	8	19

*n = total count of applied codes in the in-depth analyses of the sub-questions. Count in cells indicate frequency of cross-coding.

The cross-tabulations between other types of policies and assessment type (Table 4.17), region (Table 4.18), education level (Table 4.19) and sampling approach (Table 4.20) did not reveal any noteworthy differences, largely due to the small numbers in the cells as a consequence of the relatively large number of sub-categories under both resource allocation and teaching and learning practices.

4.7.4 Facilitators and barriers - synthesis

Table 4.21: Facilitators, barriers and assessment type

Factors		Sub-national (state-level) (n=4)*	National (n=31)*	Regional (n=16)*	International (n=21)*
Soundness of programme	Facilitator	1	5	5	6
	Barrier	3	11	1	6
Integration into policy processes	Facilitator	1	6	7	4
	Barrier	1	3	2	1
Further analysis	Facilitator	0	5	3	3
	Barrier	2	9	2	3
Analysis to diagnose Issues	Facilitator	0	4	1	3
	Barrier	2	6	3	5
Media/public opinion	Facilitator	0	5	5	8
	Barrier	1	1	1	2
Dissemination to general public	Facilitator	1	6	4	6
	Barrier	1	6	4	3
Stakeholders receive appropriate results	Facilitator	1	7	6	5
	Barrier	1	7	3	2

*n = total count of applied codes in the in-depth analyses of the sub-questions. Count in cells indicate frequency of cross-coding.

Table 4.21 presents a cross-tabulation between types of assessment programme and whether factors related to the nature of the assessment programme itself were cited in the literature as a facilitator or a barrier to the use of data from the programme. Some key differences emerge.

For both sub-national and national assessments, the soundness (or lack thereof) of the assessment programme is more often seen as a barrier to the use of data in policy, rather than as a facilitator. However, as noted previously, the small number of studies on sub-national assessments means that this finding should be interpreted cautiously. For regional assessments, the soundness of programmes are more often seen as a facilitator while for international assessments, the quality or soundness of the programme is seen equally often as a facilitator and a barrier.

Integration into the policy process is a key facilitator to the use of data from assessment programmes for national, regional and international assessments. Similarly, media/public opinion is more often seen as a facilitator than a barrier to the use of assessment data. Dissemination activities for education reform stakeholders were more often seen as being appropriate and facilitating the use of assessment data than serving as a barrier. It is interesting to note that dissemination to the general public is seen almost equally often as a facilitator and a barrier. This may be due to the fact that when dissemination to the public served as a barrier, dissemination activities were inadequate and the information supplied was not relevant to the public's needs. Lastly, dissemination activities to the public and stakeholders may be more critical for national assessment programmes, as these activities were seen to be both facilitators and barriers.

Table 4.22: Facilitators, barriers and region

Factor		Europe (n=5)*	Africa - North of Sahara (n=4)*	Africa - South of Sahara (n=66)*	North and Central America (n=18)*	South America (n=49)*	Asia (n=2 4)*	Middle East (n=2)*	Pacific (n=3)*
Soundness of programme	Facilitator	2	1	5	2	6	3	1	0
	Barrier	0	0	4	1	9	5	0	0
Integration into policy processes	Facilitator	1	0	9	1	3	5	2	0
	Barrier	0	0	2	1	1	1	0	0
Further analysis	Facilitator	0	0	3	0	2	1	1	0
	Barrier	0	0	0	0	2	6	3	0
Analysis to diagnose issues	Facilitator	1	1	3	0	5	2	1	0
	Barrier	0	0	3	1	3	2	0	0
Media/public opinion	Facilitator	4	1	7	2	9	6	1	0
	Barrier	0	0	1	1	1	0	0	0
Disseminatio n to general public	Facilitator	2	0	6	1	6	4	2	0
	Barrier	1	0	2	3	5	3	1	0
Stakeholders receive appropriate results	Facilitator	3	2	7	3	9	4	1	0
	Barrier	0	0	1	3	5	2	0	0

*n = total count of applied codes in the in-depth analyses of the sub-questions.
Count in cells indicate frequency of cross-coding.

In Table 4.22 facilitators and barriers related to the nature of the assessment programme are cross-tabulated with regions. Looking at factors across regions, the soundness of the assessment programme was slightly more often seen as a barrier than a facilitator, while integration into policy processes and media/public opinion were more often seen as facilitators to the use of assessment data in policy-making. Examining factors within regions, particularly for Africa - South of Sahara, South America and Asia, because of the high frequency of assessment programmes coded in these regions, the media and public opinion were key facilitators to the use of data in education policy, especially in South America. The assessment programme's integration in policy processes served as a facilitator in both Africa - South of Sahara and Asia. Also, the appropriate dissemination of information to education reform stakeholders helped to facilitate impact on policy in these two regions.

Very few records in this review include the Pacific. Consequently there is a significant knowledge gap in this region of factors that facilitate or hinder the use of assessment data in education policy-making.

5. Summary of results

This systematic review examined the available literature on the use of data from large-scale assessments in education policy-making in developing countries. An extensive literature search initially identified close to 1,500 records. Ultimately, 54 studies were included in this review, with 73 countries represented out of 151 countries that fall under the review's definition of an economically developing country.

The body of available literature suggests that an understanding of the ways in which large-scale assessment programmes have influenced and can influence policy is currently largely confined to certain geographic areas. National assessments in countries in South America and regional assessments from Africa - South of Sahara contributed the most to the material in this review. Meanwhile, this review shows that less is known about the ways in which assessments have been used for policy-making in developing countries in Asia and especially in the Pacific.

This review identified a substantial number of studies that referred to an assessment programme being carried out in a developing country, with recommendations made based on its findings. As they did not provide information on whether or not the recommendations were adopted, this group of studies was excluded from the review and separately analysed as 'academic papers'. A comparison between this group of academic papers and the included studies was undertaken in terms of geographical coverage. Results showed that academic papers covered 16 developing countries in addition to the 73 countries that were covered in the included review material, but there was no evidence on whether or not the academic papers in which they featured ever influenced policy-making or practice.

According to a recent paper, about 120 developing countries have taken part in either a national or regional assessment programme (Simons 2012). However, only 65 countries are represented in this review as having had a study report on the use of a regional or national assessment programme in policy-making. Thus, this systematic review has identified examples of a link between regional or national assessment programmes and policy-making in slightly more than half of the developing countries that have undertaken these assessments.

Considering the body of literature included in this review, large-scale assessments were able to be characterised along several dimensions including: type; country/region; sampling approach and education level. These characteristics, then, allowed a detailed description of large-scale assessment programmes for which independent evidence of links to policy-making were found in terms of the material retrieved for this review. Together, this enabled the following description of large-scale assessments which have had a link to policy-making:

- Just under half of the assessment programmes in the review were national in coverage, followed by one-third international programmes, while approximately one-fifth were regional assessment programmes and only a few were sub-national assessment programmes.
- Of the regional assessment programmes, the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) featured most often, followed by the Latin American Laboratory for Assessment of the Quality of Education (LLCE/SERCE) and the Programme d'analyse des systèmes éducatifs

de la Conférence des ministres de l'éducation des pays ayant le français en partage (PASEC).

- Of the international assessments, the Programme for International Student Assessment (PISA) featured most often, followed by the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS).
- Africa-South of Sahara was the region that featured the most in the review, followed by South America. These more frequent occurrences were probably a consequence of the SACMEQ/PASEC and LLECE assessment programmes in these regions.
- Most large-scale assessment programmes operated at both the primary and secondary education levels, or solely at the primary education level. The smallest proportion of assessments was undertaken at the secondary education level only.
- Most of the assessments used a representative sample rather than a census for data collection, probably due to the cost-prohibitive nature of undertaking a census.
- Large-scale assessment programmes were most often intended to measure and ensure educational quality. Assessment programmes were less often used for the policy goals of equity, accountability and leverage for specific education matters.
- Relative to the total number of references for each region, quality as an explicit policy goal for the use of large-scale assessments occurred more frequently in South America and Asia than in the developing countries in Africa, Europe, North and Central America and the Pacific region.
- Considering the link between assessment programme data and the policy process, and regardless of whether an assessment was sub-national, national, regional or international in type, assessment data were used slightly more often in three stages of the policy cycle, namely (i) policy agenda setting, (ii) policy implementation and (iii) policy monitoring and evaluation than for the stage of policy formulation. In other words, large-scale assessments were used had a slightly lower impact on the ways in which analytical and political options and strategies for education policies were constructed than on other types of policy activities.
- Large-scale assessments conducting a census of a target population almost equally mentioned the policy goals of quality, equity and accountability, while assessments that used a representative sampling approach were connected more to quality as a policy goal than to equity or accountability. Considering the impact on the policy process, assessments that used a representative sampling approach had relatively more impact than census-based assessments on the policy formulation stage. The most frequently mentioned resource allocation policies were aimed at increasing teacher quality and teaching materials through in-service professional development, improved teacher preparation, and textbook reform. To a lesser extent, large-scale assessment data were used for resource allocation policies related to changes in education funding.
- Impact on teaching and learning practice policies was observed less frequently in the review than impact on policies regarding resource allocations. Of these policies, the most targeted ones focused on learning processes by way of student-oriented pedagogy and in-class learning strategies. A key *de novo* teaching and learning practices policy emerged, i.e. the development and dissemination of targeted teacher and leadership guides to assist with improving practice and leadership.

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

- The most frequent education policies resulting from the use of assessment data were system-level policies regarding (i) curriculum standards and reform, (ii) performance standards, and (iii) assessment policies.
- Notably, slightly less than half of all records described an assessment programme that impacted upon curriculum standards whereas two-thirds of records referring to international assessments reported an impact on curriculum standards.
- The most common facilitators for assessment data to be used in policy-making, regardless of the type of assessment programme, were media and public opinion, appropriate dissemination to stakeholders, the soundness of the assessment programme and the programme's integration into policy processes.
- *De novo* facilitating factors were further identified from the reviewed materials, and these highlighted the importance of funding as well as the continuity and stability of the assessment agency and programme.
- The most commonly noted barriers to the use of assessment data were the level of quality or soundness of the assessment programme and the inability to undertake further, in-depth analyses of the data.
- The quality of the assessment programme was more often seen as a facilitator to the use of regional assessment data, while the lack of quality was more often seen as a barrier to the use of sub-national and national assessments. For international assessment programmes, the quality of the assessment programme emerged as a facilitator as well as a barrier in equal proportions.
- Appropriateness and extent of dissemination activities targeted at the general public were equally likely to be reported as a facilitator and a barrier to the use of assessment data.
- Records that explicitly noted no impact on the policy process outlined barriers to the use of assessment data, which were thematically grouped as problems relating to (i) assessment programme and analyses, (ii) financial constraints, (iii) weak assessment bodies and fragmented government agencies, and (iv) low technical capacity of assessment staff.
- Few studies examined the role of factors outside the assessment programmes themselves (i.e. the nature of the education, social and political systems) in acting as facilitators or barriers to the use of large-scale assessment data in policy-making.

6. Implications of the review

6.1 Implications for intended audience of the review

The primary stakeholders of this review are those involved in the planning and funding of large-scale assessments and using data from them in developing countries. The results of this review are intended to guide the use of assessment data and participation in assessment programmes.

6.2 Strengths and limitations

Almost two-thirds of all developing countries have participated in a national, regional or international assessment programme, though little is known about the use of large-scale assessment data in policy-making. This review is the first study which systematically examines a wide body of literature to synthesise the available evidence on the relationship between these assessment programmes and education policy in developing countries. However, the review faced several challenges as rigorous policy analysis for large-scale assessments is scarce. Therefore, the literature that linked education policies to large-scale assessments and the policy process were few. Despite these limitations, this review has been able to move beyond existing knowledge of large-scale assessment activities in developing countries to map associations between factors associated with assessment programmes and policy-making in developing countries.

This review did not exclude literature based on appraised quality. As such, evidence from literature appraised as being of 'low quality' was included in the synthesis of evidence. The authors initially intended to conduct sensitivity analyses to assess whether or not the conclusions applied to a body of high-quality literature. As there was not a large body of literature included in the review overall, further reducing the number of records would have led to an insufficient number of cases with many empty cells in the results tables. Therefore, while it may have been of interest to conduct sensitivity analyses, it was considered to be more of an academic exercise that would do little to contribute to the meaningfulness of the study's results.

6.3 Opportunities for further research

Based on the findings of the systematic review, a number of suggestions regarding further research are briefly outlined here.

Several information gaps have been identified in this review, which may guide future studies that examine the link between large-scale assessments and policy-making in developing countries. Concerning assessment programmes, little material could be found about large-scale sub-national assessments that have been undertaken in decentralised education systems, such as India and Brazil, and their links to policy-making. Usually in systems such as these, sub-national states or districts may be comparable in size with national education systems in other countries and warrant further investigation.

Regionally, less is known about the impact of large-scale assessment programmes in countries within Asia, the Pacific and the Middle East and North Africa. As noted, literature from Africa - South of Sahara and South America principally contributed to the richness of the review, in part due to well-established regional assessment programmes. The absence of a regional assessment covering developing countries in the Asia-Pacific region is noteworthy. Given the relative success of SACMEQ in

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

terms of its impact on policy-making, it would be worthwhile to investigate in more detail what exactly contributed to the effectiveness of this assessment programme.

Results presented in this review have shed some light on the facilitators and barriers to the potential of large-scale assessments for impacting on policy-making. However, it seems desirable to examine further potential ways of increasing the policy-assessment link, perhaps by organising events attended by both policy-makers and people who develop and implement large-scale assessment, and researching possibly in a more qualitative way the obstacles to greater exchange, cross-fertilisation and cooperation.

Considering facilitators and barriers at a system level, factors beyond the nature of the assessment programme which were associated with the wider education and political systems were not found to be key drivers for the use of assessment data in policy-making. These factors included, for example, the effectiveness of the education system, political sensitivities and conflict, the strength of civil society and public discourse. The lack of association between these wider factors and impact on policy could be an artefact from the included literature as the rigour of the policy analysis in the available literature varied widely. Future research and policy analysis may aim to examine the relationship between system-level factors as barriers and facilitators of education policy-making in developing countries.

Finally, it would be of great interest to do the same study for developed countries. While some studies have been commissioned by the organisations that are undertaking large-scale assessments (Breakspear 2012, Gilmore 2005) an independent systematic review would provide stronger evidence of the existence or absence of links between large-scale assessments and education policy-making.

7. References

*Publicly available records, included for detailed synthesis are demarcated by **

* Abdul-Hamid H, Abu-Lebdeh KM, Patrinos HA (2011) *Assessment testing can be used to inform policy decisions: the case of Jordan*. PolicyResearch Working Paper 5890. Washington, DC: World Bank Human Development Network.

Abulibdeh K, Abdelsamad M (2008) Jordan: analysis of differences in mathematics and science achievement according to student gender, school location, school authority, and school resources. In: Lietz P, Wagemaker H, Neuschmidt O, Henke J (eds) *Educational issues in the Middle East North Africa region*. Hamburg: International Association for the Evaluation of Educational Achievement, pages 57-76.

* Arancibia V (1997) *Los sistemas de medición y evaluación de la calidad de la educación*. Laboratorio Latinomaeicano de Evaluación de la calidad de la educación. Documentos. UNESCO. <http://unesdoc.unesco.org/images/0018/001836/183651s.pdf>

Assessment and Evaluation Center (2006) *National report on the results of TIMSS 2003*. Ramallah: Ministry of Education and Higher Education.

Bantick C (2011) My school mute on bad teachers. *The Australian*, 5 March 2011. www.theaustralian.com.au/news/opinion/my-school-mute-on-bad-teachers/story-e6frg6zo-1226016103319 (accessed 2 August 2011).

Barnett-Page E, Thomas J (2009) Methods for the synthesis of qualitative research: a critical review. *BMC Medical Research Methodology* 9(59): 1-12.

Benavot A, Tanner E (2007) The growth of national learning assessments in the world, 1995-2006. In: *Education for All by 2015: will we make it? EFA global monitoring report 2008*. Paris: UNESCO, pages 1-17.

* Benveniste L (2000) Student assessment as political construction: the case of Uruguay. *Education Policy Analysis Archives* 8(32): 1-40.

* Bernard JM, Michaelowa K (2006) Managing the impact of PASEC projects in francophone sub-Saharan Africa. In: Ross K, Genevois IJ (eds) *Cross-national studies of the quality of education: planning their design and managing their Impact*. Paris: UNESCO, International Institute for Educational Planning (IIEP), pages 229-240.

Berry R, Adamson B (2011) Assessment reform past, present and future. In: Berry R, Adamson B (eds) *Assessment reform in education. Education in the Asia-Pacific region: issues, concerns and prospects*. New York: Springer, pages 3-14.

* Bloch G (2009) *The education roadmap in South Africa*. Policy Brief Number 9. Brighton, UK: Consortium for Research on Educational Access, Transitions and Equity (CREATE).

Braslavsky C (2005) The history of education and the contemporary challenge of quality education for all. *Prospects* 35(4): 385-399.

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

Braun H, Kanjee A (2006) Using assessment to improve education in developing nations. In: Braun H, Kanjee A, Bettinger E, Kremer M (eds) *Improving education through assessment, innovation, and evaluation*. Cambridge, MA: American Academy of Arts and Sciences, pages 1-46.

* Breakspear S (2012) *The policy impact of PISA: an exploration of the normative effects of international benchmarking in school system performance*. OECD Education Working Papers No. 71. OECD Publishing.
<http://dx.doi.org/10.1787/5k9fdfqffr28-en>

Bridgman P, Davis G (2004) *The Australian policy handbook*, 3rd edn. Sydney: Allen and Unwin.

Campbell C, Levin B (2009) Using data to support educational improvement. *Educational Assessment, Evaluation and Accountability* 21: 47-65.

* Cariola L, Covacevich C, Gubler J, Lagos E, Ortiz M (2011) Chilean participation in IEA studies. In: Papanastasiou C, Plomb T, Papanastasiou E (eds) *IEA 1958-2008: 50 years of experiences and memories. Volume 1*. Nicosia: Cultural Center of the Kykkos Monastery, pages 373-387.

Carroll C, Booth A, Cooper K (2011) A worked example of 'best fit' framework synthesis: a systematic review of views concerning the taking of some potential chemo preventive agents. *BMC Medical Research Methodology* 11(29): 1-9.

* Castro MH, (2010) Brazil: creating an enabling environment for an effective assessment system. Slides presented at: *Second Russian Education Aid for Development (READ) Global Conference*, Nairobi, Kenya. http://siteresources.worldbank.org/EDUCATION/Resources/2782001256666213814/READ_GCII_17_MHelena_English.pdf

* Castro MH, Tiezzi S (2004) The reform of secondary education and the implementation of ENEM in Brazil. *Oxford Studies in Comparative Education* 13: 89-113.

* Chinapah V (2000) *With Africa for Africa: towards quality education for all*. 1999 MLA Project. Pretoria: Human Sciences Research Council/UNESCO/UNICEF.

* CONFEMEN (2001) *Stratégies pour une refondation réussie des systèmes éducatifs: document de réflexion et d'orientation*. Dakar: CONFEMEN.

* Crespo M, Soares JF, de Mello e Souza A (2000) The Brazilian national evaluation system of basic education: context, process and impact. *Studies in Educational Evaluation* 26(2): 105-125.

Crotty M (1998) *The foundations of social research: meaning and perspective in the research process*. Singapore: Allen and Unwin.

Darling-Hammond L (2004) Standards, accountability and school reform. *Teachers College Record* 106(6): 1047-1085.

Davies D (2004) Is evidence-based government possible? Jerry Lee Lecture, presented at: *Fourth Annual Campbell Collaboration Colloquium*, Washington, DC.

Delandshere G (1997) History of educational research. In: Keeves JP (ed) *Educational research, methodology and measurement: an international handbook*, 2nd edn. Oxford: Pergamon Press, pages 8-16.

*Department of Basic Education (2011) *Delivery agreement for outcome 1: improved quality of basic education*. South Africa: Ministry of Education.
www.info.gov.za/view/DownloadFileAction?id=135271

Dixon-Woods M (2011) Using framework-based synthesis for conducting reviews of qualitative studies. *BMC Medicine* 9(39): 1-2.

* Elley WB (2005) How TIMSS-R contributed to education in eighteen developing countries. *Prospects* 35(2): 199-212.

* Ferrer G (2006) *Educational assessment systems in Latin America: current practice and future challenges*. Washington, DC: Partnership for Educational Revitalization in the Americas (PREAL).

* Fiske EB (2000) *Education for all: status and trends, 2000. Assessing learning achievement*. Paris: UNESCO.

Forster M (2001) *A policy maker's guide to systemwide assessment programs*. Camberwell: Australian Council for Educational Research.
http://research.acer.edu.au/policy_makers_guides/2

* Gilmore A (2005) *The impact of PIRLS (2001) and TIMSS (2003) in low and middle-income countries: an evaluation of the value of World Bank support for international surveys of reading literacy (PIRLS) and mathematics and science (TIMSS)*. Amsterdam: International Association for the Evaluation of Educational Achievement.

Goldstein H, Thomas SM (2008) Assessment in education: principles, policy and practice. *International Comparative Studies in Achievement*, Special Issue 15(3): 211-320.

* Greaney V, Kellaghan T (2008) *Assessing national achievement levels in education. National assessments of educational achievement. Volume 1*. Washington, DC: World Bank.

* Gutiérrez DMA, Vázquez F G (2008) *Resultados nacionales: segundo estudio regional comparativo y explicativo 2006 (SERCE)*. Mexico, DF: Instituto Nacional para la Evaluación de la Educación (INEE).

* Gutiérrez DMA, Vázquez FG (2010) *México en PISA 2009*. Mexico, DF: Instituto Nacional para la Evaluación de la Educación (INEE).

* Gvirtz S, Larripa S (2004) National evaluation system in Argentina: problematic present and uncertain future. *Assessment in Education* 11(3): 349-364.

Haddad WD (1995) *Education policy-planning process: an applied framework*. Paris: UNESCO, International Institute for Educational Planning (IIEP).

Hanushek EA (2003) The failure of input-based schooling policies. *The Economic Journal* 113:F64-F98.

Hanushek EA, Kimko DD (2000) Schooling, labor-force quality and the growth of nations. *The American Economic Review* 90: 1184-1208.

Hardy I, Boyle C (2011) My school? Critiquing the abstraction and quantification of education. *Asia-Pacific Journal of Teacher Education* 39(3): 211-222.

Hattie J (2009) *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Husén T, Kogan M (1984) *Educational research and policy: how do they relate?* Oxford: Pergamon Press.

JB I (2011a) *The Joanna Briggs Institute reviewers' manual: 2011 edition*. Adelaide: Joanna Briggs Institute.

JB I (2011b) *SUMARI user manual: version 5. System for the unified management, assessment and review of information*. Adelaide: Joanna Briggs Institute.

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

Kamens DH, Benavot A (2011) National, regional and international learning assessments: trends among developing countries, 1960-2009. *Globalisation, Societies and Education* 9(2): 285-300.

Kamens DH, McNeely CL (2009) Globalization and the growth of international educational testing and national assessment. *Comparative Education Review* 54(1): 5-25.

* Kanjee A, Acana S (2010) Uganda: creating an enabling environment for an effective assessment system. Slides presented at: *Second Russian Education Aid for Development (READ) Global Conference*, Nairobi, Kenya. http://siteresources.worldbank.org/EDUCATION/Resources/2782001256666213814/READ_GCII_16_AKanjee_English.pdf

Keeves JP (1995) *The world of school learning: selected key findings from 35 years of IEA research*. The Hague: International Association for the Evaluation of Educational Achievement.

Kellaghan T, Greaney V, Murray TS (2009) *Using the results of a national assessment of educational achievement. National assessments of educational achievement. Volume 5*. Washington, DC: World Bank.

* Kellaghan T, Bethell G, Ross J (2011) *National and international assessments of student achievement*. Guidance Note: a DFID Practice Paper. London: Department for International Development.

*Khaniya T, Williams JH (2004) Necessary but not sufficient: challenges to (implicit) theories of educational change: reform in Nepal's primary education system. *International Journal of Educational Development* 24(3): 315-328.

Kos J, Nugroho D, Lietz P (2010) *Quality of education in madrasah*. Submitted by the Australian Council for Educational Research (ACER) to Contractor Strategic Advisory Services (CSAS).

Krueger AB (2002) Understanding the magnitude and effect of class size on student achievement. In: Mishel L, Rothstein R (eds) *The class size debate*. Washington, DC: Economic Policy Institute, pages 7-35.

*Leste A (2005) Streaming in Seychelles: from SACMEQ research to policy reform. Paper presented at *SACMEQ Research Conference*. Paris: SACMEQ.

Lietz P, Wagemaker H, Neuschmidt O, Hencke J (2008) *Educational issues in the Middle East North Africa region: outcomes of the IEA Arab region training seminar series 2006/2007*. The Hague: IEA.

Lubisi RC, Murphy RJL (2002) Assessment in South African schools. *Assessment in Education: Principles, Policy and Practice* 9(2): 255-268.

* Maligalig DS, Albert JR (2008) *Measures for assessing basic education in the Philippines*. Discussion Paper Series No. 2008-16. Manila: Philippine Institute for Development Studies.

* Marchelli PS (2010) Expansion and quality of primary and secondary education in Brazil. *Cadernos de Pesquisa* 40(140): 561-585.

* Marcondes MI (1999) Teacher education in Brazil. *Journal of Education for Teaching* 25(3): 203-213.

Marston G, Watts R (2003) Tampering with the evidence: a critical appraisal of evidence-based policy-making. *The Drawing Board: an Australian Review of Public Affairs* 3(3): 143-163.

- * Martínez AS (2007) *Casos de estudio de reacciones a PISA 2000 y 2003: Alemania, Polonia, Brasil y otros*. Mexico, DF: Fundación IDEA (Implementación, Diseño, Evaluación y Análisis de Políticas Públicas).
- * Meckes L, Carrasco R (2010) Two decades of SIMCE: an overview of the national assessment system in Chile. *Assessment in Education: Principles, Policy and Practice* 17(2):233-248.
- * Mesa V, Gómez P, Cheah U (in press) Impact of international studies of student achievement on mathematics teaching and learning. In: Clements MA, Bishop A, Keitel C, Kilpatrick J, Leung F(eds) *Third international handbook of mathematics education*. New York: Springer.
- * Ministry of Education (2004) *The development of education. National report of Pakistan*. Islamabad: Government of Pakistan.
www.ibe.unesco.org/International/ICE47/English/Natreps/reports/pakistan.pdf
- * Mizala A, Urquiola M (2007) *School markets: the impact of information approximating schools' effectiveness*. NBER Working Paper No. 13676. Cambridge, MA: National Bureau of Economic Research.
- * Mourshed M, Chijioke C, Barber M (2010) *Education: how the world's most improved school systems keep getting better*. London: McKinsey and Company. Available at: https://mckinseysociety.com/downloads/reports/education/how-the-worlds-most-improved-school-systems-keep-getting-better_download-version_final.pdf, last accessed 8 May 2013.
- * Mukhopadhyay R, Sriprakash A (2010) Global frameworks, local contingencies: policy translations and education development in India. *Compare: a Journal of Comparative and International Education* 41(3): 311-326.
- Mullis IVS, Martin MO, Ruddock GJ, O'Sullivan CY, Arora A, Erberber E (2005) *TIMSS 2007 assessment frameworks*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College.
- * Ni Y, Li Q, Li X, Zhang Z (2011) Influence of curriculum reform: an analysis of student mathematics achievement in mainland China. *International Journal of Educational Research* 50:100-116.
- * Nielsen DH (2006) *From schooling access to learning outcomes, an unfinished agenda: an evaluation of World Bank support to primary education*. Washington, DC: World Bank.
- * Nzomo J, Makuwa D (2006). How can countries move from dissemination, and then to policy reform? (Case studies from Kenya and Namibia). Chapter 10 in: Ross K, Genevois IJ (eds) *Cross-national studies of the quality of education*. Paris: UNESCO, International Institute for Educational Planning (IIEP), pages 213-228.
- Oliver SR, Rees RW, Clarke-Jones L, Milne R, Oakley AR, Gabbay J, Stein K, Buchanan P, Gyte G (2008) A multidimensional conceptual framework for analysing public involvement in health services research. *Health Expectations* 11(1): 72-84.
- OECD (2009a) *Effective teaching and learning environments: first results from TALIS*. Paris: OECD Publishing.
- OECD (2009b) *PISA 2009 assessment framework: key competencies in reading, mathematics and science*. Paris: OECD Publishing.
- OECD (2009c) *The policies that inspired PISA - and the policies that PISA inspired: 2009 report on the evaluation of the policy impact of PISA*. Paris: OECD Publishing.
- OECD (2010) *Education at a glance 2011: OECD indicators*. Paris: OECD Publishing.

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

* Pacific Resources for Education and Learning (1999) *Pacific Language Use in the Schools (PLUS) study. Progress report*. Washington, DC: Office of Educational Research and Improvement.

Pawson R (2002) Evidence-based policy: in search of a method. *Evaluation* 8(2): 157-181.

Pollard A, Court J (2005) *How civil society organisations use evidence to influence policy processes: a literature review*. Overseas Development Institute Working Paper 249. London: Overseas Development Institute.

Popham WJ (1999) Why standardized test scores don't measure educational quality. *Educational Leadership* 56(6): 8-15.

Postlethwaite TN (1984) Research and policy-making in education: some possible links. In: Husén T, Kogan M (eds) *Educational research and policy: how do they relate?* Oxford: Pergamon Press, pages 195-206).

Postlethwaite TN, Kellaghan T (2008) *National assessments of educational achievement*. Education Policy Series 9. Paris: UNESCO, International Institute for Educational Planning (IIEP) and Brussels: International Academy of Education (IAE).

* Ramirez MJ (2010) Chile: creating an enabling environment for an effective assessment system. Slides presented at: *Second Russian Education Aid for Development (READ) Global Conference*, Nairobi, Kenya. http://siteresources.worldbank.org/EDUCATION/Resources/2782001256666213814/READ_GCII_13_MRamirez_English.pdf

* Ravela P (2005) A formative approach to national assessments: the case of Uruguay. *Prospects: Quarterly Review of Comparative Education* 35(1): 21-43.

Ravitch D (2010) Why I changed my mind. *The Nation*, 14 June 2010. www.thenation.com/article/why-i-changed-my-mind (accessed 1 August 2011).

* Reddy V (2005) Cross-national achievement studies: learning from South Africa's participation in the Trends in International Mathematics and Science Study (TIMSS). *Compare: a Journal of Comparative and International Education* 35(1): 63-77.

* Reimers F (2007) Civic education when democracy is in flux: the impact of empirical research on policy and practice in Latin America. *International Journal of Citizenship Teaching and Learning* 3(2): 5-21.

* Ruto, S. (2010) Uwezo: shifting the focus from enrollment to learning. Slides presented at: *Second Russian Education Aid for Development (READ) Global Conference*, Nairobi, Kenya. http://siteresources.worldbank.org/EDUCATION/Resources/2782001256666213814/READ_GCII_6_SRuto_English.pdf

Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS (1996) Evidence based medicine: what it is and what it isn't. *BMJ* 312: 71-72 (editorial).

* Saito M, vanCappelle F (2010) Monitoring the quality of education: exploration of concept, methodology, and the link between research and policy. In: Wiseman AW (ed) *The impact of international achievement studies on national education policymaking*. International Perspectives on Education and Society, Volume 13. Bingley, UK: Emerald Group, pages 3-34.

* Sarva Shiksha Abhiyan (2011) *Thirteenth joint review mission, 17th to 31st January 2011. Aide memoire*. <http://ssa.nic.in/>

* Shamatov D, Sainazarov K (2010) The impact of standardized testing on education quality in Kyrgyzstan: the case of the Program for International Student Assessment (PISA) 2006. In: Wiseman AW (ed) *The impact of international achievement studies on national education policymaking*. International Perspectives on Education and Society, Volume 13. Bingley, UK: Emerald Group, pages 145-179.

Simons KA (2012) *Learning Metrics Task Force discussion paper #1: multi-country assessments of learning*. Washington, DC: Brookings Institution.

* Smith WJ, Ngoma-Maema WY (2003). Education for all in South Africa: developing a national system for quality assurance. *Comparative Education* 39(3): 345-365.

Solesbury W (2001) *Evidence based policy: whence it came and where it's going*. Working Paper 1. Swindon: ESRC UK Centre for Evidence Based Policy and Practice.

* Souza PR (2005) *Linking empirical research to basic (k12) education policies: the Brazilian case 1995-2002*. Presented at: 'Research for Results in Education' Global Conference on Education Research in Developed and Transition Countries, Prague, Czech Republic. <http://www.preal.org/>

Sunderman G, Orfield G (2008) *Domesticating a revolution: No Child Left Behind reforms and state administrative response*. Cambridge, MA: Harvard Education Publishing Group.

Sutcliffe S, Court J (2005) *Evidence-based policymaking: what is it? How does it work? What relevance for developing countries?* London: Overseas Development Institute.

Topsfield J (2011) Labor may have won education wars, but it is a pyrrhic victory. *The Age*, 12 March 2011. www.theage.com.au/opinion/politics/labor-may-have-won-education-wars-but-its-a-pyrrhic-victory-20110311-1br8q.html (accessed 5 August 2011).

Triccoa A, Tetzlaff J, Sampson M, Fergusson D, Cogoa E, Horsley T, Moher D (2008) Few systematic reviews exist documenting the extent of bias: a systematic review. *Journal of Clinical Epidemiology* 61: 422-434.

UNESCO (2008) *Education for All by 2015: will we make it? EFA global monitoring report 2008*. Paris: UNESCO.

Wiseman AW (2010) The uses of evidence for educational policymaking: global contexts and international trends. *Review of Research in Education* 34(1): 1-24.

Woessmann L (2000) *Schooling resources, educational institutions, and student performance: the international evidence*. Working Paper 983. Kiel: Kiel Institute for World Economics.

* Wolff L, Navarro JC, González P (eds) (2005) *Private education and public policy in Latin America*. Washington, DC: Partnership for Educational Revitalization in the Americas (PREAL).

Young E, Quinn L (2002) *Writing effective public policy papers: a guide for policy advisers in central and eastern Europe*. Budapest: Local Government and Public Service Reform Initiative, Open Society Institute.

Appendices

Appendix 1.1: Authorship of this report

Authors (in alphabetical order):

Ms Maura Best, ACER

Ms Pat Knight, ACER

Dr Petra Lietz, ACER

Dr Craig Lockwood, JBI

Ms Dita Nugroho, ACER

Ms Mollie Tobin, ACER

Review group membership

Dr Petra Lietz (ACER), Dr Michele Lonsdale (ACER), Ms Dita Nugroho (ACER), Ms Pat Knight (ACER), Ms Mollie Tobin (ACER), Dr John Ainley (ACER), Ms Maura Best (ACER), Dr Craig Lockwood (JBI)

Details of Peer Review Group membership

Peer review expertise	Peer review member	Affiliation
Systematic review	Nominated by EPPI-Centre	
Policy area	Nominated by AusAID	
Topic area interest	David Rutkowski, PhD	Assistant research scientist, Center for Evaluation and Education Policy, Indiana University

Advisory Group membership

Lucrecia Santibanez, PhD, Economist, Center for Latin American Social Policy, RAND Corporation

Maurice Robson, PhD, Chief of Education Section, Pakistan, United Nations Children's Fund

This report should be cited as:

Best M, Knight P, Lietz P, Lockwood C, Nugroho D, Tobin M (2013) The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries. Final report. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

ISBN: 978-1-907345-57-9

Contact details

Dr Petra Lietz
Level 3, 97 Pirie Street
AdelaideSA5000

Tel: 08 7129 4493
Fax: 08 7129 4446
lietz@acer.edu.au
www.acer.edu.au

Conflict of interest

The authors have no financial interests to disclose. We have some experience in the conduct of national assessment studies in Indonesia (Kos et al. 2010) and Mexico. The authors' home organisation, ACER, has extensive involvement in the implementation of international assessment programmes, as well as more recent consultative roles for national assessment programmes in developing countries. This may lead to a perception of positive bias on the impact of assessment programmes. We have mitigated this potential by nominating David Rutkowski as a peer reviewer who has questioned the rationale, purpose and intentions of assessment programmes.

Acknowledgements

We wish to acknowledge the feedback provided by the advisory group and thank them for their help. We also wish to thank the individuals that responded to our solicitation for grey literature.

Appendix 1.2: Classification of developing countries

Table A1.1 Classification of developing countries¹ (n=151)

Developing countries in alphabetical order	Developing countries by region
Afghanistan	Africa - North of Sahara
Albania	Algeria
Algeria	Egypt
Angola	Libya
Anguilla	Morocco
Antigua and Barbuda	Tunisia
Argentina	Africa - South of Sahara
Armenia	Angola
Azerbaijan	Benin
Bangladesh	Botswana
Barbados	Burkina Faso
Belarus	Burundi
Belize	Cameroon
Benin	Cape Verde
Bhutan	Central African Republic
Bolivia	Chad
Bosnia and Herzegovina	Comoros
Botswana	Congo, Republic
Brazil	Congo, Democratic Republic
Burkina Faso	Cote d'Ivoire
Burma (Myanmar)	Djibouti
Burundi	Equatorial Guinea
Cambodia	Eritrea
Cameroon	Ethiopia
Cape Verde	Gabon

Appendix 1.2: Classification of developing countries

Central African Republic	Gambia, The
Chad	Ghana
Chile	Guinea
China (excluding Hong Kong)	Guinea-Bissau
Colombia	Kenya
Comoros	Lesotho
Congo, Democratic Republic	Liberia
Congo, Republic	Madagascar
Cook Islands	Malawi
Costa Rica	Mali
Côte d'Ivoire	Mauritania
Croatia	Mauritius
Cuba	Mozambique
Djibouti	Namibia
Dominica	Niger
Dominican Republic	Nigeria
East Timor (Timor Leste)	Rwanda
Ecuador	St. Helena
Egypt	São Tomé and Príncipe
El Salvador	Senegal
Equatorial Guinea	Seychelles
Eritrea	Sierra Leone
Ethiopia	Somalia
Fiji	South Africa
Gabon	South Sudan
Gambia, The	Sudan
Georgia	Swaziland
Ghana	Tanzania

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

Grenada	Togo
Guatemala	Uganda
Guinea	Zambia
Guinea-Bissau	Zimbabwe
Guyana	Asia
Haiti	Afghanistan
Honduras	Bangladesh
India	Bhutan
Indonesia	Burma
Iran, Islamic Republic	Cambodia
Iraq	China (excl. Hong Kong)
Jamaica	East Timor (Timor Leste)
Jordan	India
Kazakhstan	Indonesia
Kenya	Kazakhstan
Kiribati	Korea, Democratic Republic
Korea, Democratic Republic	Kyrgyz Republic
Kyrgyz Republic	Laos, People's Democratic Republic
Laos, People's Democratic Republic	Malaysia
Lebanon	Maldives
Lesotho	Mongolia
Liberia	Nepal
Libya	Pakistan
Macedonia, Former Yugoslav Republic	Philippines
Madagascar	Sri Lanka
Malawi	Tajikistan
Malaysia	Thailand
Maldives	Turkmenistan

Mali	Uzbekistan
Marshall Islands	Vietnam
Mauritania	Europe
Mauritius	Albania
Mayotte	Armenia
Mexico	Azerbaijan
Micronesia, Federated States of	Belarus
Moldova	Bosnia and Herzegovina
Mongolia	Georgia
Montserrat	Kosovo
Morocco	Macedonia, Former Yugoslav Republic
Mozambique	Moldova
Namibia	Serbia and Montenegro
Nauru	Turkey
Nepal	Ukraine
Nicaragua	Middle East
Niger	Iran
Nigeria	Iraq
Niue	Jordan
Oman	Lebanon
Pakistan	Palestine (Territories Administered by the Palestinian Authority)
Palau	
Panama	Syria
Papua New Guinea	Yemen
Paraguay	Pacific
Peru	Cook Islands
Philippines	Micronesia, Federated States
Rwanda	Fiji
Samoa	Kiribati
São Tomé and Príncipe	Marshall Islands
Senegal	Nauru
Serbia and Montenegro	Niue
Seychelles	Palau Islands

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

Sierra Leone	Papua New Guinea
Solomon Islands	Samoa
Somalia	Solomon Islands
South Africa	Tokelau
Sri Lanka	Tonga
St. Helena	Tuvalu
St. Kitts and Nevis	Vanuatu
St. Lucia	Wallis and Futuna
St. Vincent and the Grenadines	North and Central America
Sudan	Anguilla
Suriname	Antigua and Barbuda
Swaziland	Belize
Syria	Costa Rica
Tajikistan	Cuba
Tanzania	Dominica
Palestine (Territories Administered by the Palestinian Authority)	Dominican Republic
Thailand	El Salvador
Togo	Grenada
Tokelau	Guatemala
Tonga	Haiti
Trinidad and Tobago	Honduras
Tunisia	Jamaica
Turkey	Mexico
Turkmenistan	Montserrat
Tuvalu	Nicaragua
Uganda	Panama
Ukraine	St. Kitts and Nevis
Uruguay	St. Lucia
Uzbekistan	St. Vincent and the Grenadines
Vanuatu	South America
Venezuela	Argentina
Vietnam	Bolivia
Wallis and Futuna	Brazil
Yemen, Republic	Chile
Zambia	Colombia
Zimbabwe	Ecuador

Appendix 1.2: Classification of developing countries

Guyana
Paraguay
Peru
Suriname
Uruguay
Venezuela

¹ AusAID (2009) *List of developing countries: as declared by the Minister for Foreign Affairs*. Available at: www.aisaid.gov.au/ngos/Documents/list-developing-countries.doc last accessed 8 May 2013.

Appendix 2.1: Exclusion criteria in screening

Code in review software	Criteria
Excl published prior to 1990	Studies published earlier than 1990 are excluded
Excl not developing country	Developing countries listed in Appendix 1.2
Excl audience not education policy-makers or practitioners	Audience for paper must be either education policy-makers, or practitioners. Otherwise, paper is excluded
Excl not primary or secondary education setting	Settings must be primary or secondary school environments
Excl not a standardised assessment of academic achievement	Standardisation requires consistency in test design, content, administration and scoring to ensure comparability of the results across students and schools
Excl single-schools or studies not representative of an administrative area	Studies must be state, national, or multi-national (regional) in scope
Excl no data for impact on or recommendations for teaching and learning practices	Impact being evidence of policy-making processes for resource allocation and/or teaching or learning practices
Excl for duplicate record	Duplicate records are excluded
Include for detailed assessment	Study included if exclusion criteria not applied

Appendix 2.2: Database keywords and descriptors

SCOPUS - SEARCH TERMS

52 records retrieved using the following search terms limited to items published between 1990 and 2011

Subject area

Social sciences and humanities

Article title, abstract, keywords fields

(NB Scopus does not have a controlled vocabulary to describe subjects so there is no subject field. Instead the terms were searched across the article title, abstract and keyword fields)

Student assessment	International assessment
National assessment	International student assessment
Student performance	National student assessment
High-stakes test*	Achievement test*
Standardised test*	Standardized test*
Educational test*	Examinations*
National exam*	

101,034 records retrieved

Article title, abstract keywords fields

Elementary education	Primary education
Secondary education	Middle school education
Elementary school education	Primary school education
Secondary school education	Middle school education

5,116 records retrieved

Article title, abstract keywords fields

Afghanistan or Albania or Algeria or Angola or Anguilla or Antigua or Argentina or Armenia or Azerbaijan or Bahamas or Bahrain or Bangladesh or Barbados or Barbuda or Belarus or Belize or Benin or Bhutan or Bolivia or Bosnia or Botswana or Brazil or Brunei or Bulgaria or Burkina or Burma or Burundi or Cambodia or Cameroon or “Cape Verde” or “Central African Republic” or Chad or Chile or China or Colombia or Comoros or Congo or “Cook Islands” or “Costa Rica” or “Cote d’Ivoire” or “Cote d Ivoire” or Croatia or Cuba or Darussalam or Djibouti or Dominica* or “East Timor” or Ecuador or Egypt or “El Salvador” or Eritrea or Estonia or Ethiopia or Fiji or Futuna or Gabon or Gambia or Georgia or Ghana or Grenada or Grenadines or Guatemala or Guinea* or Guyana or Haiti or Herzegovina or Honduras or Hungary or India or Indonesia or Iran or Iraq or Jamaica or Jordon or Kazakhstan or Kenya or Kiribati or

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

Korea or Kosovo or Kuwait or Kyrgyz or Laos or Latvia or Lebanon or Lesotho or Liberia or Libya or Lithuania or Macedonia or Madagascar or Malawi or Malaysia or Maldives or Mali or “Marshall Islands” or Mauritania or Mauritius or Mayotte or Mexico or Micronesia or Moldova or Mongolia or Montserrat or Montenegro or Morocco or Mozambique or Myanmar or Namibia or Nauru or Nepal or Nevis or Nicaragua or Niger or Nigeria or Niue or Oman or Pakistan or Palau or Panama or “Papua New Guinea” or Paraguay or Peru or Philippines or Poland or Principe or Qatar or Romania or Russia or Rwanda or Samoa or “Sao Tome” or “Saudi Arabia” or Senegal or Serbia or Seychelles or “Sierra Leone” or “Solomon Islands” or Somalia or “South Africa” or “Sri Lanka” or “St Helena” or “St Kitts” or “St Lucia” or “St Vincent” or Sudan or Suriname or Swaziland or Syria or Tajikistan or Tanzania or Palestine or Thailand or “Timor Leste” or Togo or Tokelau or Tonga or Trinidad or Tobago or Tunisia or Turkey or Turkmenistan or Tuvalu or Uganda or Ukraine or “United Arab Emirates” or Uruguay or Uzbekistan or Vanuatu or Venezuela or Vietnam or Wallis or Yemen or Zambia or Zimbabwe

108,845 records retrieved

Prepared by Pat Knight, 7 September 2011

EDUCATION RESOURCES INFORMATION CENTER (ERIC) - SEARCH TERMS

550 records retrieved using the following search terms limited to items published between 1990 and 2011

Subject field

Student evaluation Educational testing

Achievement tests Standardised tests

Testing programs National standards

Testing National competency tests

National exam* (*n.b. this is not a recognised subject descriptor so the term was searched across the entire record*)

39,147 records retrieved

Subject field

Elementary education

Elementary secondary education

Secondary education

200,512 records retrieved

Subject field

(*NB this subject descriptor was excluded from the search using the Boolean operator NOT*)

College entrance

3,648

Identifier field

(NB identifier field was used as these country names are not subject descriptors, but identifiers)

Afghanistan or Albania or Algeria or Angola or Anguilla or Antigua or Argentina or Armenia or Azerbaijan or Bahamas or Bahrain or Bangladesh or Barbados or Barbuda or Belarus or Belize or Benin or Bhutan or Bolivia or Bosnia or Botswana or Brazil or Brunei or Bulgaria or Burkina or Burma or Burundi or Cambodia or Cameroon or “Cape Verde” or “Central African Republic” or Chad or Chile or China or Colombia or Comoros or Congo or “Cook Islands” or “Costa Rica” or “Cote d’Ivoire” or “Cote d Ivoire” or Croatia or Cuba or Darussalam or Djibouti or Dominica* or “East Timor” or Ecuador or Egypt or “El Salvador” or Eritrea or Estonia or Ethiopia or Fiji or Futuna or Gabon or Gambia or Georgia or Ghana or Grenada or Grenadines or Guatemala or Guinea* or Guyana or Haiti or Herzegovina or Honduras or Hungary or India or Indonesia or Iran or Iraq or Jamaica or Jordon or Kazakhstan or Kenya or Kiribati or Korea or Kosovo or Kuwait or Kyrgyz or Laos or Latvia or Lebanon or Lesotho or Liberia or Libya or Lithuania or Macedonia or Madagascar or Malawi or Malaysia or Maldives or Mali or “Marshall Islands” or Mauritania or Mauritius or Mayotte or Mexico or Micronesia or Moldova or Mongolia or Montserrat or Montenegro or Morocco or Mozambique or Myanmar or Namibia or Nauru or Nepal or Nevis or Nicaragua or Niger or Nigeria or Niue or Oman or Pakistan or Palau or Panama or “Papua New Guinea” or Paraguay or Peru or Philippines or Poland or Principe or Qatar or Romania or Russia or Rwanda or Samoa or “Sao Tome” or “Saudi Arabia” or Senegal or Serbia or Seychelles or “Sierra Leone” or “Solomon Islands” or Somalia or “South Africa” or “Sri Lanka” or “St Helena” or “St Kitts” or “St Lucia” or “St Vincent” or Sudan or Suriname or Swaziland or Syria or Tajikistan or Tanzania or Palestine or Thailand or “Timor Leste” or Togo or Tokelau or Tonga or Trinidad or Tobago or Tunisia or Turkey or Turkmenistan or Tuvalu or Uganda or Ukraine or “United Arab Emirates” or Uruguay or Uzbekistan or Vanuatu or Venezuela or Vietnam or Wallis or Yemen or Zambia or Zimbabwe

33,294 records retrieved

(NB ERIC uses the subject descriptor ‘Developing Nations’ but this was not used in the search as it likely that there would be inconsistencies between ERIC indexers as to what is deemed to be a developing country. Instead, the countries identified by AusAID, the IMF and the World Bank were each searched)

Prepared by Pat Knight, 29 August 2011

EDUCATION RESEARCH COMPLETE (ERC) - SEARCH TERMS

140 records retrieved using the following search terms limited to items published between 1990 and 2011

Subject field

Educational tests and measurements

High-stakes tests

Academic achievement testing

Academic achievement

Standardized tests

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

Competency based educational tests

Examinations

Testing

National competency based educational tests

National exam* (*NB this is not a recognised subject descriptor so the term was searched across the entire record*)

Subject field

Elementary education

Education elementary

Primary education

Secondary education

Education secondary

Middle school education

Subject field

(NB ERC has a Geographic Terms field, but the Subject field was searched as the same search in the Geographic Terms field yielded over 7,000 fewer results)

Afghanistan or Albania or Algeria or Angola or Anguilla or Antigua or Argentina or Armenia or Azerbaijan or Bahamas or Bahrain or Bangladesh or Barbados or Barbuda or Belarus or Belize or Benin or Bhutan or Bolivia or Bosnia or Botswana or Brazil or Brunei or Bulgaria or Burkina or Burma or Burundi or Cambodia or Cameroon or "Cape Verde" or "Central African Republic" or Chad or Chile or China or Colombia or Comoros or Congo or "Cook Islands" or "Costa Rica" or "Cote d'Ivoire" or "Cote d'Ivoire" or Croatia or Cuba or Darussalam or Djibouti or Dominica* or "East Timor" or Ecuador or Egypt or "El Salvador" or Eritrea or Estonia or Ethiopia or Fiji or Futuna or Gabon or Gambia or Georgia or Ghana or Grenada or Grenadines or Guatemala or Guinea* or Guyana or Haiti or Herzegovina or Honduras or Hungary or India or Indonesia or Iran or Iraq or Jamaica or Jordon or Kazakhstan or Kenya or Kiribati or Korea or Kosovo or Kuwait or Kyrgyz or Laos or Latvia or Lebanon or Lesotho or Liberia or Libya or Lithuania or Macedonia or Madagascar or Malawi or Malaysia or Maldives or Mali or "Marshall Islands" or Mauritania or Mauritius or Mayotte or Mexico or Micronesia or Moldova or Mongolia or Montserrat or Montenegro or Morocco or Mozambique or Myanmar or Namibia or Nauru or Nepal or Nevis or Nicaragua or Niger or Nigeria or Niue or Oman or Pakistan or Palau or Panama or "Papua New Guinea" or Paraguay or Peru or Philippines or Poland or Principe or Qatar or Romania or Russia or Rwanda or Samoa or "Sao Tome" or "Saudi Arabia" or Senegal or Serbia or Seychelles or "Sierra Leone" or "Solomon Islands" or Somalia or "South Africa" or "Sri Lanka" or "St Helena" or "St Kitts" or "St Lucia" or "St Vincent" or Sudan or Suriname or Swaziland or Syria or Tajikistan or Tanzania or Palestine or Thailand or "Timor Leste" or Togo or Tokelau or Tonga or Trinidad or Tobago or Tunisia or Turkey or Turkmenistan or Tuvalu or Uganda or Ukraine or "United Arab Emirates" or Uruguay or Uzbekistan or Vanuatu or Venezuela or Vietnam or Wallis or Yemen or Zambia or Zimbabwe

Prepared by Pat Knight, 7 September 2011

BRITISH EDUCATION INDEX (BEI) - SEARCH TERMS

81 records retrieved using the following search terms limited to items published between 1990 and 2011

Subject field

Student evaluation Educational testing
Achievement tests Standardised tests
Standardised tests Testing programmes
Testing programs National standards
Testing National competency tests

Examinations

National exam* (*NB this is not a recognised subject descriptor so the term was searched across the entire record*)

6,693 records retrieved

Subject field

Primary education
Primary secondary education
Secondary education

45,002 records retrieved

Countries and regions field

(NB countries and regions field was used as these country names are not subject descriptors)

Afghanistan or Albania or Algeria or Angola or Anguilla or Antigua or Argentina or Armenia or Azerbaijan or Bahamas or Bahrain or Bangladesh or Barbados or Barbuda or Belarus or Belize or Benin or Bhutan or Bolivia or Bosnia or Botswana or Brazil or Brunei or Bulgaria or Burkina or Burma or Burundi or Cambodia or Cameroon or “Cape Verde” or “Central African Republic” or Chad or Chile or China or Colombia or Comoros or Congo or “Cook Islands” or “Costa Rica” or “Cote d Ivore” or Croatia or Cuba or Darussalam or Djibouti or Dominica\$ or “East Timor” or Ecuador or Egypt or “El Salvador” or Eritrea or Estonia or Ethiopia or Fiji or Futuna or Gabon or Gambia or Georgia or Ghana or Grenada or Grenadines or Guatemala or Guinea\$ or Guyana or Haiti or Herzegovina or Honduras or Hungary or India or Indonesia or Iran or Iraq or Jamaica or Jordon or Kazakhstan or Kenya or Kiribati or Korea or Kosovo or Kuwait or Kyrgyz or Laos or Latvia or Lebanon or Lesotho or Liberia or Libya or Lithuania or Macedonia or Madagascar or Malawi or Malaysia or Maldives or Mali or “Marshall Islands” or Mauritania or Mauritius or Mayotte or Mexico or Micronesia or Moldova or Mongolia or Montserrat or Montenegro or Morocco or Mozambique or Myanmar or Namibia or Nauru or Nepal or Nevis or Nicaragua or Niger or Nigeria or Niue or Oman or Pakistan or Palau or Panama or “Papua New Guinea” or Paraguay or Peru or Philippines or Poland or Principe or Qatar or Romania or Russia or Rwanda or Samoa or “Sao Tome” or “Saudi Arabia” or Senegal or Serbia or Seychelles or “Sierra Leone” or “Solomon Islands” or Somalia or “South Africa” or “Sri Lanka” or “St Helena” or “St Kitts” or “St Lucia” or “St Vincent” or Sudan or Suriname or Swaziland or Syria or Tajikistan or Tanzania or Palestine or Thailand or

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

“Timor Leste” or Togo or Tokelau or Tonga or Trinidad or Tobago or Tunisia or Turkey or Turkmenistan or Tuvalu or Uganda or Ukraine or “United Arab Emirates” or Uruguay or Uzbekistan or Vanuatu or Venezuela or Vietnam or Wallis or Yemen or Zambia or Zimbabwe

8,373 records retrieved

Prepared by Pat Knight, 8 September 2011

AUSTRALIAN EDUCATION INDEX (aka A+ EDUCATION) - SEARCH TERMS

73 records retrieved using the following search terms limited to items published between 1990 and 2011

Subject field

Student assessment Educational testing

Achievement tests Testing programs

National standards Standardised tests

Testing National competency tests

National exam* (*NB this is not a recognised subject descriptor so the term was searched across the entire record*)

7,307 records retrieved

Minor subject field(*NB Australian Education Index uses this field to indicate level of education and, in cases where the record is research-based, the research method[s] used*)

Primary education

Primary secondary education

Secondary education

32,910 records retrieved

Subject field

(*NB these terms were excluded from the search by the use of the Boolean operator NOT*)

College entrance examinations

University entrance examinations

Screening tests

215 records retrieved

Geographic location field

Afghanistan or Albania or Algeria or Angola or Anguilla or Antigua or Argentina or Armenia or Azerbaijan or Bahamas or Bahrain or Bangladesh or Barbados or Barbuda

or Belarus or Belize or Benin or Bhutan or Bolivia or Bosnia or Botswana or Brazil or Brunei or Bulgaria or Burkina or Burma or Burundi or Cambodia or Cameroon or “Cape Verde” or “Central African Republic” or Chad or Chile or China or Colombia or Comoros or Congo or “Cook Islands” or “Costa Rica” or “Cote d’Ivoire” or “Cote d Ivoire” or Croatia or Cuba or Darussalam or Djibouti or Dominica* or “East Timor” or Ecuador or Egypt or “El Salvador” or Eritrea or Estonia or Ethiopia or Fiji or Futuna or Gabon or Gambia or Georgia or Ghana or Grenada or Grenadines or Guatemala or Guinea* or Guyana or Haiti or Herzegovina or Honduras or Hungary or India or Indonesia or Iran or Iraq or Jamaica or Jordon or Kazakhstan or Kenya or Kiribati or Korea or Kosovo or Kuwait or Kyrgyz or Laos or Latvia or Lebanon or Lesotho or Liberia or Libya or Lithuania or Macedonia or Madagascar or Malawi or Malaysia or Maldives or Mali or “Marshall Islands” or Mauritania or Mauritius or Mayotte or Mexico or Micronesia or Moldova or Mongolia or Montserrat or Montenegro or Morocco or Mozambique or Myanmar or Namibia or Nauru or Nepal or Nevis or Nicaragua or Niger or Nigeria or Niue or Oman or Pakistan or Palau or Panama or “Papua New Guinea” or Paraguay or Peru or Philippines or Poland or Principe or Qatar or Romania or Russia or Rwanda or Samoa or “Sao Tome” or “Saudi Arabia” or Senegal or Serbia or Seychelles or “Sierra Leone” or “Solomon Islands” or Somalia or “South Africa” or “Sri Lanka” or “St Helena” or “St Kitts” or “St Lucia” or “St Vincent” or Sudan or Suriname or Swaziland or Syria or Tajikistan or Tanzania or Palestine or Thailand or “Timor Leste” or Togo or Tokelau or Tonga or Trinidad or Tobago or Tunisia or Turkey or Turkmenistan or Tuvalu or Uganda or Ukraine or “United Arab Emirates” or Uruguay or Uzbekistan or Vanuatu or Venezuela or Vietnam or Wallis or Yemen or Zambia or Zimbabwe

5,827 records retrieved

Prepared by Pat Knight, 6 September 2011

ELDIS, ASIA-STUDIES ONLINE, BRITISH LIBRARY FOR DEVELOPMENT STUDIES (BLDS) and JOLIS - SEARCH TERMS

ELDIS	ASIA STUDIES ONLINE
www.eldis.org/	www.asia-studies.com
<ul style="list-style-type: none"> • Search restricted to the topic ‘Education’ • Number in brackets indicate the number of records retrieved 	<ul style="list-style-type: none"> • Boolean search used • “w/4” searches for the first term within 4 words, in either order of the second term • - is the truncation symbol
student assessment school (1)	"student assessment" w/4 school- (3)
educational test school (5)	"educational test-" w/4 school- (1)
achievement test school (9)	"achievement test-" w/4 school- (9)
national standards school (2)	"national standards" w/4 school- (6)
standardised tests school (0)	"standardised test-" w/4 school- (0)
standardized tests school (0)	"standardized test-" w/4 school- (1)
standardised tests (0)	"national competency tests" w/4 school- (0)
standardized tests (0)	testing w/4 school- (71)
standardised test (0)	"national exam-" w/4 school- (4)
	"student evaluation" w/4 school- (0)

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

<p>standardized test (0) national competency test school (1) testing school (55) national examination school (3) student evaluation school (0) student evaluation (1) examinations school (0) examinations (0) exams (0) educational tests school (5) international assessment school (9) international student assessment school (3) national assessment school (16) national student assessment school (1) student performance school (1)</p>	<p>exam- w/4 school- (48) “international assessment” w/4 school- (0) “international student assessment” w/4 school- (0) “national assessment” w/4 school- (1) “international student assessment” w/4 school- (0) “student performance” w/4 school- (18)</p>
<p>BLDS Library Catalogue http://blds.ids.ac.uk/search-the-collection <ul style="list-style-type: none"> Limited to publication dates 1990-2011 SU=testing and SU=education (8) SU=examinations and SU=education (21) SU=measurement and SU=education (21)</p>	<p>JOLIS Library Catalogue http://jolis.worldbankimflib.org/e-nljolis.htm <ul style="list-style-type: none"> Searches limited to the subject field and publication date >1989 educational tests and measurements - developing countries (2) educational evaluation - developing countries (5) academic achievement AND developing countries (15) achievement tests (4)</p>

Appendix 2.3: Data extraction tool with keywords

Review-specific keywords

1. At what level is the assessment programme implemented?	<ul style="list-style-type: none"> 1.1 Sub-national 1.2 National 1.3 Regional 1.4 International
2. In what country(ies) and region(s) was the assessment programme implemented?	<ul style="list-style-type: none"> 2.1 Countries listed in Appendix 1.2 2.2 Asia 2.3 Europe 2.4 Middle East 2.5 Pacific 2.6 North and Central America 2.7 Africa -North of Sahara 2.8 South America 2.9 Africa -South of Sahara
3. What is the sampling strategy of the assessment programme?	<ul style="list-style-type: none"> 3.1 Population/census 3.2 Representative sample
4. What level of education does the programme assess?	<ul style="list-style-type: none"> 4.1 Primary 4.2 Secondary 4.3 Both primary and secondary
5. What stage of the policy process is influenced by assessment as described in the study?	<ul style="list-style-type: none"> 5.1 Agenda setting 5.2 Policy formulation 5.3 Policy implementation 5.4 Monitoring and policy evaluation 5.5 No impact on policy process 5.6 Other
6. What is the goal of the use of assessment described in the study?	<ul style="list-style-type: none"> 6.1 Measure and ensure quality 6.2 Measure and ensure equity 6.3 Accountability 6.4 Leverage 6.6 Other
7. What policies resulting from the use of assessment data in policy-making are described in the study?	<p>Resource allocation</p> <ul style="list-style-type: none"> 7.1 Instructional materials 7.2 School supplies

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

	<p>7.3 Equipment</p> <p>7.4 Facilities</p> <p>7.5 School feeding/meals</p> <p>7.6 Class size/ratios</p> <p>7.7 Instructional time/school hours</p> <p>7.8 Teacher preparation</p> <p>7.9 Teacher recruitment and retention</p> <p>7.10 In-service professional development</p> <p>7.11 Funding formula</p> <p>7.12 Decision-making authority</p> <p>7.13 Other</p> <p>Teaching and learning practices</p> <p>7.14 Classroom management and discipline</p> <p>7.15 In-class learning strategies</p> <p>7.16 Student-oriented pedagogy</p> <p>7.17 Enhanced learning activities</p> <p>7.18 Staff collaboration/mentoring</p> <p>7.19 Student-teacher relationships</p> <p>7.20 Teacher attitudes</p> <p>7.21 Organisation of instructional/study time</p> <p>7.22 Additional classes</p> <p>7.23 Extra-curricular activities</p> <p>7.24 Motivation and future plans</p> <p>7.25 Tracking/streaming policy</p> <p>7.26 Other</p> <p>System level</p> <p>7.27 Assessment policy</p> <p>7.28 Curriculum standards</p> <p>7.29 Performance standards</p> <p>7.30 Community/parent engagement policy</p> <p>7.31 Accountability policy</p> <p>7.32 Other</p>
<p>8. What facilitators or barriers to the use of assessment data are described in the study?</p>	<p>8.1 Soundness of programme</p> <p>8.2 Integration into policy processes</p> <p>8.3 Policy-maker involvement in assessment programme</p> <p>8.4 Further analysis</p>

	<p>8.5 Analysis to diagnose issues</p> <p>8.6 Timing of results</p> <p>8.7 Stakeholders receive appropriate results</p> <p>8.8 Value of assessment findings and research</p> <p>8.9 Effectiveness of education system</p> <p>8.10 Teacher unions</p> <p>8.11 Decision-making channels to schools</p> <p>8.12 Political sensitivities</p> <p>8.13 Decentralisation and openness</p> <p>8.14 Strength of structures for civil society</p> <p>8.15 Freedom of public discourse</p> <p>8.16 Accountability systems</p> <p>8.17 Role of external agencies</p> <p>8.18 Media/public opinion</p> <p>8.19 Dissemination to general public</p> <p>8.20 Other</p>
--	---

Appendix 2.4: Synthesis table

Policy			Facilitators	Barriers					Quality	Equity		Accountability	} <i>Goals</i>	
Resource allocation			U	V					←-----→	Leverage and 'Other'				
Teaching and learning			W	X										
Other			Y	Z					Agenda setting	Policy formulation		Policy implementation	Monitoring and evaluation	<i>Policy stages</i>
Regions: Asia, Europe, Middle East, North and Central America, Africa - North of Sahara, South America, Africa - South of Sahara	National	Primary	Census	A	B	C	D	E	NO POLICY IMPACT					
	Regional	Secondary	Sample	F	G	H	I	J						
				K	L	M	N	O						
	International	Both primary and secondary		P	Q	R	S	T						

Regions *Assessment types* *Levels of education* *Sampling approach*

**Appendix 3.1: Description of ‘academic papers with possible policy suggestions’
(n=78)**

Table A3.1 Keyworded region and countries in academic papers

Region and country	Frequency of applied keywords
Asia	27
Bangladesh	2
Bhutan	1
Cambodia	1
China (excluding Hong Kong)	2
India	5
Indonesia	3
Laos	1
Malaysia	2
Nepal	1
Philippines	1
Thailand	4
Timor-Leste	1
Vietnam	3
Europe	10
Albania	1
Macedonia	1
Moldova	1
Turkey	7
Middle East	1
Iran	1
North and Central America	12
Belize	1
Cuba	1
Honduras	1
Jamaica	1
Mexico	8
Pacific	4
Micronesia, Federated States	2
Papua New Guinea	2
South America	26

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

Argentina	8
Bolivia	1
Brazil	6
Chile	5
Colombia	2
Peru	2
Uruguay	2
Africa - South of Sahara	117
Botswana	6
Burkina Faso	1
Cameroon	1
Chad	1
Côte d'Ivoire	1
Guinea	1
Kenya	13
Lesotho	5
Madagascar	1
Malawi	8
Mali	1
Mauritania	1
Mauritius	7
Mozambique	5
Namibia	9
Niger	2
Nigeria	1
Senegal	1
Seychelles	5
South Africa	16
Swaziland	5
Tanzania	7
Togo	2
Uganda	5
Zambia	7
Zimbabwe	5

Table A3.2 Keyworded level of assessment programme in academic papers

Level of assessment programme implemented	Frequency of applied keywords
Sub-national (state level)	4
National	47
Regional	21
International	10

Table A3.3 Keyworded sampling strategy of assessment programme in academic papers

Sampling strategy of assessment programme	Frequency of applied keywords
Population/census	15
Representative sample	50

Table A3.4 Keyworded level of education of assessment programme in academic papers

Level of education in assessment programme	Frequency of applied keywords
Primary education	41
Secondary education	23
Both primary and secondary education	13

Table A3.5 Keyworded resource allocation policy suggestions in academic papers

Resource allocation policy suggestion	Frequency of applied keywords
In-service professional development	14
Instructional materials	14
Teacher preparation	14
Funding formula	10
Facilities	9
Class size	8
Decision-making authority	8
Equipment	7
Teacher recruitment and retention	7
Other	7
Instructional time/school hours	4
School supplies	4
School feeding/meals	3

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

Table A3.6 Keyworded teaching and learning policy suggestions in academic papers

Teaching and learning policy suggestion	Frequency of applied keywords
Other	9
In-class learning strategies	7
Student oriented pedagogy	7
Teacher attitudes	6
Enhanced learning activities	4
Classroom management and discipline	3
Motivation and future plans	3
Additional classes	2
Extra-curricular activities	2
Student-teacher relationships	2
Organisation of instructional/study time	1
Staff collaboration/mentoring	1

Table A3.7 Keyworded policy suggestions in academic papers

All other policy suggestion	Frequency of applied keywords
Other	18
Curriculum standards	9
Assessment policy	7
Community/parent engagement policy	7
Performance standards	7
Accountability policy	6
Tracking/streaming policy	3

Appendix 3.2: Synthesis table of included records for detailed analysis (n=50)

Source/date	Region	Document type	Quality		Assessment type			Policy stages						Policy impacts		
			High	Low	National	Regional	International	Agenda setting	Policy formulation	Policy implementation	Monitoring and evaluation	No impact on policy process	Other	Resource allocation	Teaching and learning	System level
ABDUL-HAMID et al. (2011)	Middle East	Working paper		✓	✓		✓	✓		✓				✓	✓	✓
ARANCIBIA (1997)	South America; North and Central America	UNESCO report	✓		✓		✓		✓	✓		✓			✓	✓
BENVENISTE (2000)	South America	Policy analysis	✓		✓		✓		✓	✓	✓			✓		✓
BERNARD and MICHAELWA (2006)	Africa - South of Sahara		✓			✓			✓	✓	✓			✓	✓	✓
BLOCH (2009)	Africa - South of Sahara	Policy brief		✓	✓		✓					✓				✓
BREAKSPEAR (2012)	Europe; North and Central America; South America; Asia	Report		✓			✓	✓	✓		✓	✓			✓	✓
CARIOLA et al. (2011)	South America	Book chapter	✓		✓	✓	✓			✓	✓					✓
CASTRO and TIEZZI (2004)	South America	Journal article	✓		✓		✓				✓				✓	
CASTRO (2010)	South America	Conference proceedings		✓	✓			✓	✓		✓	✓				✓
CHINAPAH (2000)	Africa - North of Sahara, South of Sahara	Book - MLA project		✓		✓					✓					✓
CONFEMEN (2001)	Africa- South of Sahara	Report		✓		✓		✓	✓		✓			✓		✓

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

Source/date	Region	Document type	Quality		Assessment type			Policy stages						Policy impacts			
			High	Low	National	Regional	International	Agenda setting	Policy formulation	Policy implementation	Monitoring and evaluation	No impact on policy process	Other	Resource allocation	Teaching and learning	System level	
CRESPO et al. (2000)	South America	Journal article	✓		✓				✓	✓					✓		✓
Department of Basic Education (2011)	Africa - South of Sahara	Report- delivery agreement		✓		✓	✓	✓									✓
ELLEY (2005)	Europe; Asia	Journal article	✓				✓	✓		✓	✓				✓	✓	✓
FERRER (2006)	North and Central America; South America	Book	✓		✓	✓	✓	✓		✓	✓	✓			✓	✓	✓
FISKE (2000)	Africa South of Sahara; South America; Asia	Report	✓		✓		✓	✓	✓	✓	✓	✓			✓	✓	✓
GILMORE (2005)	Europe; Asia; Africa - North of Sahara, South of Sahara; South America	Evaluation report	✓				✓		✓	✓	✓	✓				✓	✓
GREANEY and KELLAGHAN (2008)	Africa- South of Sahara; South America; Asia	Book		✓	✓							✓			✓		✓
GUTIERREZ and VAZQUEZ (2008)	North and Central America	Book	✓			✓		Unknown/not applicable								✓	
GUTIERREZ and VAZQUEZ(2010)	North and Central America	Report	✓			✓			✓	✓					✓	✓	✓
GVIRTZ and LARRIPA(2004)	South America	Journal article	✓		✓						✓				✓		
KANJEE and ACANA	Africa - South of	Conference		✓	✓			Unknown/not applicable								✓	

Appendix 3.2: Synthesis table of included records for detailed analysis (n=50)

Source/date	Region	Document type	Quality		Assessment type			Policy stages						Policy impacts			
			High	Low	National	Regional	International	Agenda setting	Policy formulation	Policy implementation	Monitoring and evaluation	No impact on policy process	Other	Resource allocation	Teaching and learning	System level	
(2010)	Sahara	proceedings															
KELLAGHAN, BETHELL and ROSS (2011)	Africa - South of Sahara; South America; Asia	Guidance note/ Practice paper	✓		✓			✓	✓	✓				✓	✓	✓	
KHANIYA and WILLIAMS (2004)	Asia	Article	✓		✓					✓							✓
LESTE (2005)	Africa - South of Sahara	Conference proceedings	✓			✓		✓	✓						✓		
MALIGALIG and ALBERT (2008)	Asia	Report		✓	✓		✓				✓						✓
MARCHELLI (2010)	South America	Journal article		✓	✓			✓	✓	✓							✓
MARCONDES (1999)	South America	Journal article	✓		✓					✓				✓			✓
MARTINEZ (2007)	South America		✓				✓		✓		✓	✓	✓	✓	✓	✓	✓
MECKES and CARRASCO (2010)	South America	Journal article		✓	✓			✓	✓	✓	✓			✓	✓	✓	✓
MESA et al. (in press)	Asia; South America	Book chapter	✓				✓			✓		✓					✓
MINISTRY of EDUCATION (2004)	Asia	Report		✓	✓				✓		✓						✓
MIZALA and URQUIOLA (2007)	South America	Working paper	✓		✓								✓				✓
MOURSHED et al. (2010)	Europe; Africa - South of Sahara; Asia; South America; Middle East	Report		✓			✓			✓							✓

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

Source/date	Region	Document type	Quality		Assessment type			Policy stages						Policy impacts		
			High	Low	National	Regional	International	Agenda setting	Policy formulation	Policy implementation	Monitoring and evaluation	No impact on policy process	Other	Resource allocation	Teaching and learning	System level
MUKHOPADHYAY and SRIPRAKASH (2011)	Asia	Journal article	✓										✓	✓	✓	✓
NI et al. (2011)	Asia	Journal article		✓			✓			✓					✓	✓
NIELSON (2006)	Africa - South of Sahara; South America; Asia	Evaluation report	✓		Unknown/not applicable				✓			✓				✓
NZOMO and MAKUWA (2006)	Africa - South of Sahara	Book chapter		✓										✓	✓	✓
Pacific Resources for Education and Learning(1999)	Pacific	Report		✓		✓		✓								✓
RAMIREZ (2010)	South America	Conference proceedings		✓	✓			Unknown/not applicable								✓
RAVELA (2005)	South America	Journal article	✓		✓			✓		✓	✓			✓	✓	✓
REDDY (2005)	Africa - South of Sahara	Journal article	✓				✓			✓		✓		✓		
REIMERS (2007)	South America	Journal article	✓				✓	✓	✓		✓		✓			✓
RUTO (2010)	Africa - South of Sahara	Conference proceedings		✓		✓		✓								✓
SAITO and VANCAPELLE (2010)	Asia; Africa - South of Sahara	Book chapter	✓		✓	✓		✓		✓	✓	✓		✓		✓
SARVA SHIKSHA SABHIYAN (2011)	Asia	Report	✓		✓							✓	✓			

Appendix 3.2: Synthesis table of included records for detailed analysis (n=50)

Source/date	Region	Document type	Quality		Assessment type			Policy stages						Policy impacts			
			High	Low	National	Regional	International	Agenda setting	Policy formulation	Policy implementation	Monitoring and evaluation	No impact on policy process	Other	Resource allocation	Teaching and learning	System level	
SHAMATOV and SAINAZAROV (2010)	Asia	Book chapter	✓				✓		✓	✓					✓		✓
SMITH and NGOMA-MAEMA (2003)	Africa - South of Sahara	Journal chapter		✓	✓			Unknown/not applicable								✓	
SOUZA (2005)	South America	Conference proceedings		✓	✓		✓	✓	✓	✓	✓			✓			✓
WOLFF et al. (2005)	South America; North and Central America	Book		✓	✓	✓		✓	✓		✓	✓					✓

Appendix 4.1: Characteristics of assessment programmes in education policy-making

Table A4.1 Assessment programme by region and country

Region and country	Frequency of applied keywords	Percent of total mentions*
Asia	24	14%
Bhutan	1	7.1
<u>China (excluding Hong Kong)</u>	1	7.2
India	5	7.3
Indonesia	3	7.4
Kyrgyz, Republic	1	7.5
Malaysia	2	7.6
Nepal	1	7.7
Pakistan	2	7.8
Philippines	2	7.9
Sri Lanka	2	7.10
Thailand	1	7.11
Vietnam	3	7.12
Europe	5	3%
Armenia	1	7.13
Macedonia	2	7.14
Turkey	2	7.15
Middle East	2	1%
Jordan	2	7.16
North and Central America	18	11%
Costa Rica	2	7.17
Cuba	1	7.18
Dominican Republic	2	7.19
El Salvador	2	7.20

Appendix 4.1: Characteristics of assessment programmes in education policy-making

Guatemala	2	7.21
Honduras	2	7.22
Mexico	5	7.23
Nicaragua	1	7.24
Panama	1	7.25
Africa - North of Sahara	4	2%
Morocco	2	7.26
Tunisia	2	7.27
Pacific	3	2%
Marshall Islands	1	7.28
Micronesia, Federated States of	1	7.29
Palau	1	7.30
South America	49	29%
Argentina	4	7.31
Bolivia	2	7.32
Brazil	11	7%
Chile	13	8%
Colombia	6	7.33
Ecuador	1	7.34
Paraguay	2	7.35
Peru	3	7.36
Uruguay	4	7.37
Venezuela	3	7.38
Africa - South of Sahara	66	39%
Benin	1	7.39
Botswana	2	7.40
Burkina Faso	2	7.41
Cameroon	1	7.42
Central African Republic	1	7.43

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

Chad	1	7.44
Comoros	1	7.45
Congo, Democratic Republic	1	7.46
Côte d'Ivoire	1	7.47
Djibouti	1	7.48
Ethiopia	1	7.49
Gambia	1	7.50
Ghana	1	7.51
Guinea	3	7.52
Guinea-Bissau	1	7.53
Kenya	3	7.54
Madagascar	5	7.55
Malawi	2	7.56
Mali	3	7.57
Mauritania	1	7.58
Mauritius	1	7.59
Namibia	1	7.60
Niger	3	7.61
Nigeria	1	7.62
Senegal	4	7.63
Seychelles	1	7.64
South Africa	8	5%
Sudan	1	7.65
Tanzania	1	7.66
Togo	3	7.67
Uganda	6	7.68
Zambia	2	7.69
Zanzibar	1	7.70

*Results rounded to the nearest percent.

Table A4.2 Type of assessment programme

Type of assessment programme	Frequency of applied keywords	Percent of total mentions*
Sub-national	4	6%
National	31	43%
Regional	16	22%
SACMEQ		
LLECE/SERCE		
PASEC		
PREL (Pacific Resources for Education and Learning)		
International	21	29%
PISA		
TIMSS		
PIRLS		
MLA (UNICEF Monitoring Learning Achievement)		
IAEP (International Assessment of Educational Progress)		
TIMSS-R (TIMSS - Repeat)		
IEA CIVED (Civic Education Study)		
IEA ICCS (International Civic and Citizenship Study)		

*Results rounded to the nearest percent.

Table A4.3 Region and type of assessment programme

Type of assessment programme	Sub-national	National	Regional	International
Asia	3	9	1	9
Europe	0	0	0	4
Middle East	0	1	0	2
North and Central America	1	3	3	3
Africa - North of Sahara	0	0	1	1
Pacific	0	0	1	0
South	2	19	3	11

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

America				
Africa - South of Sahara	1	9	10	6

Table A4.4 Sampling approach of assessment programme

Sampling approach	Frequency of applied keywords
Population/census	17
Representative sample	43

Table A4.5 Level of education assessed

Level of education	Frequency of applied keywords
Primary education	25
Secondary education	15
Both primary and secondary education	27

Table A4.6 Level of education and type of assessment programme

Level of education	Sub-national	National	Regional	International
Primary education	3	12	12	6
Secondary education	0	8	2	12
Both primary and secondary education	2	20	7	12

Appendix 4.2: Goals and uses of assessment programmes

Goals/uses of assessment data as evidence in policy-making	Frequency of applied keywords	Percent of total mentions*
1. Measure of/used to ensure quality	43	36%
2. Measure of/used to ensure equity	25	21%
3. Ensure accountability	24	20%
4. Leverage	15	13%
5. Other:	12	10%
<ul style="list-style-type: none"> • To help inform future assessments and build technical capacity • To enable broad international comparisons • To provide inputs to be used at the local level for teachers, parents and students • To evaluate and examine the effects of pre-specified policies 		

*Results rounded to the nearest percent.

Appendix 4.3: Stages of the policy process

Stage of the policy process	Frequency of applied keywords	Percent of total mentions *
1. Agenda setting	26	21%
2. Policy formulation	23	18%
3. Policy implementation	27	21%
4. Monitoring and evaluation	31	25%
5. No impact on policy process	13	10%
6. Other:	6	5%
• Policy window		
• Policy borrowing		
• Disassociation between policy and actual school practice and implementation		

*Results rounded to the nearest percent.

Appendix 4.4: Education policies

Table A4.7 Resource allocation policies

Resource allocation policies	Frequency of applied keywords
In-service professional development	16
Instructional materials	16
Funding formula	13
Teacher preparation	10
Facilities	7
Teacher recruitment and retention	7
Decision-making authority	4
Equipment	4
Instructional time/school hours	4
School supplies	3
Class size/ratios	2
School feeding/meals	2
Total	88
Other resource allocation policy:	
<ul style="list-style-type: none"> • Financial incentives for private industry investment in public education/private sector partnerships • School health programmes • School transportation programmes • Resource standards and benchmarks for inputs • Introduction of multi-grade classrooms 	

Table A4.8 Teaching and learning practice policies

Teaching and learning practice policies	Frequency of applied keywords
Student-oriented pedagogy	6
In-class learning strategies	5
Enhanced learning activities	3
Motivation and future plans	3
Staff collaboration/mentoring	3
Additional classes	2
Organisation of instructional/study time	2
Tracking/streaming policy	2
Total	38

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

Other teaching and learning practice policy:

- Development and dissemination of teacher/principal guides for: background on curricular topics where teachers may have misconceptions; recommended pedagogic practices to target knowledge/skills assessed by ‘difficult items’; curricular lesson planning; selection of classroom texts; checklists for identified good practice/management; practical classroom investigations
- Classroom assessment frameworks
- Increased use of ICT (information and computer technology) in science instruction

Table A4.9 System-level policies

System-level policies	Frequency of applied keywords
Curriculum standards/reform	24
Assessment policy	19
Performance standards	18
Accountability policy	10
Community/parent engagement policy	8
Total	79

Other system-level policy:

- Whole-school interventions/multi-level assistance programmes for: low performing schools; low-SES schools; schools with high proportion of ethnic minorities
- Data monitoring policies: establishment of administrative data units; teacher professional development for analysis of national assessment data; teacher preparation; curriculum; focus on data analysis
- Reduction of grade repetition policies
- Targeted human resource recruitment of higher quality education managers

Appendix 4.5: Facilitators and barriers

Table A4.10 Facilitators

Facilitators	Frequency of applied keywords
Media/public opinion	15
Stakeholders receive appropriate results	14
Integration into policy processes	13
Soundness of programme	12
Dissemination to general public	11
Policy-maker involvement in assessment programme	11
Value of assessment findings and research	9
Decentralisation and openness	7
Further analysis	7
Analysis to diagnose issues	6
Role of external agencies	6
Decision-making channels to schools	5
Strength of structures for civil society	5
Teacher unions	5
Accountability systems	3
Timing of results	3
Political sensitivities	2
Effectiveness of education system	1
Other facilitators:	
<ul style="list-style-type: none"> • Autonomy, stability and legitimacy of assessment agency through proficient budget, long-term and regular assessments, single agency to buffer against political instability. • International or regional cross-country comparisons. • Publicly available databases for further research and dialogue. 	

The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries

Table A4.11 Barriers

Barriers	Frequency of applied keywords
Soundness of programme	14
Dissemination to general public	9
Further analysis	9
Analysis to diagnose issues	8
Decision-making channels to schools	8
Stakeholders receive appropriate results	8
Policy-maker involvement in assessment programme	6
Timing of results	6
Decentralisation and openness	5
Effectiveness of education system	5
Teacher unions	5
Political conflict	4
Political sensitivities	4
Accountability systems	3
Integration into policy processes	3
Role of external agencies	3
Value of assessment findings and research	3
Media/public opinion	2
Other barriers:	
<ul style="list-style-type: none"> • High quality assessment programmes (design, implementation, analysis) are cost-prohibitive • Assessment agencies not insulated from political changes and are de-legitimised • Lack of meaningful cross-country comparisons • Policy borrowing may render assessment programme meaningless for historical/cultural contexts of low- and middle-income countries 	

The authors of this report were supported by the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) and the Australian Agency for International Development.

The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) is part of the Social Science Research Unit (SSRU), Institute of Education, University of London.

Since 1993, we have been at the forefront of carrying out systematic reviews and developing review methods in social science and public policy. We are dedicated to making reliable research findings accessible to the people who need them, whether they are making policy, practice or personal decisions. We engage health and education policy makers, practitioners and service users in discussions about how researchers can make their work more relevant and how to use research findings.

Founded in 1990, the Social Science Research Unit (SSRU) is based at the Institute of Education, University of London. Our mission is to engage in and otherwise promote rigorous, ethical and participative social research as well as to support evidence-informed public policy and practice across a range of domains including education, health and welfare, guided by a concern for human rights, social justice and the development of human potential.

This research was funded by the Australian Agency for International Development (AusAID). The research was commissioned as part of a joint call for systematic reviews with the Department for International Development (DFID) and the International Initiative for Impact Evaluation (3ie). The views expressed are those of the authors and not necessarily those of the Commonwealth of Australia. The Commonwealth of Australia accepts no responsibility for any loss, damage or injury resulting from reliance on any of the information or views contained in this publication. The report was first published in 2013 by:

Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre)
Social Science Research Unit
Institute of Education, University of London
18 Woburn Square
London WC1H 0NR
Tel: +44 (0)20 7612 6397 <http://eppi.ioe.ac.uk> <http://www.ioe.ac.uk/ssru>

ISBN: 978-1-907345-57-9

Cover image © UNESCO / Justin Mott

The views expressed in this work are those of the authors and do not necessarily reflect the views of the EPPI-Centre or the funder. All errors and omissions remain those of the authors.

This document is available in a range of accessible formats including large print. Please contact the Institute of Education for assistance:
telephone: +44 (0)20 7947 9556 email: info@ioe.ac.uk