

The Impact of National and International Assessment Programmes on Educational Policy, Particularly Policies Regarding Resource Allocation and Teaching and Learning Practices in Developing Countries

Protocol written by Pat Knight, Petra Lietz, Dita Nugroho, Mollie Tobin at ACER

EPPI-Centre
Social Science Research Unit
Institute of Education
University of London

March 2012

PROTOCOL

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

This protocol should be cited as: Knight P, Lietz P, Nugroho D, Tobin M (2012) The Impact of National and International Assessment Programmes on Educational Policy, Particularly Policies Regarding Resource Allocation and Teaching and Learning Practices in Developing Countries. Protocol. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London

© 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.

Table of Contents

1.	BACKGROUND	7
1.1	Aims and rationale for current review	7
1.2	Definitional and conceptual issues	8
1.3	Theory, policy and practice background	11
1.4	Research background	18
1.5	Authors, funders, and other users of the review	19
1.6	Review questions	21
2.	METHODS USED IN THE REVIEW	21
2.1	User involvement	21
2.2	Identifying and describing studies	22
2.3	In-depth review	28
3.	REFERENCES	31

APPENDICES

Appendix 1.1: Authorship of this report

Appendix 1.2.3: Classification of developing countries

Appendix 2.2.3: Inclusion and exclusion criteria

Appendix 2.2.6: EPPI-Centre keyword sheet including review-specific keywords

Appendix 2.3.4: Preliminary synthesis table

List of abbreviations

ACER	Australian Council for Educational Research
AusAID	Australian Agency for International Development
BEI	British Education Index
DFID	Department for International Development (UK)
ERIC	Education Resources Information Center
ETS	Educational Testing Service (USA)
IEA	International Association for the Evaluation of Educational Achievement
INGO	International Nongovernmental Organisation
JBI	Joanna Briggs Institute (AUS)
LLECE	Latin American Laboratory for Assessment of the Quality of Education
NGO	Nongovernmental Organisation
OECD	Organisation for Economic Co-operation and Development
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
RTI	Research Triangle Institute (USA)
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
TALIS	Teaching and Learning International Survey
TIMSS	Trends in International Mathematics and Science Study
UK	United Kingdom
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
3ie	International Initiative for Impact Evaluation

CHAPTER ONE

1. Background

Outline of chapter

This chapter provides the background to this review. It aims to set the review within a theoretical, policy and practice context, as well as reviewing preceding relevant research into 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.

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 this review seeks to address.

1.1 Aims and rationale for current review

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

Much less is known, however, on how these assessments affect educational policy and practice in developing countries.

This review will examine 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 stems 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 & Abdelsamad, 2008; Assessment and Evaluation Center, 2006).

This review will synthesise evidence by employing a framework synthesis approach to accommodate the anticipated diverse types and quality of the literature. The use of an initial conceptual framework will effectively guide analysis to consider established evidence as well as policy considerations. At the same time, the use of a preliminary conceptual framework will also allow 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 use of 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 Educational Policy, Particularly Policies Regarding Resource Allocation and Teaching and Learning Practices in Developing Countries" are elaborated and defined below:

1.2.1 National 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 (Kamens & McNeely, 2009; Greaney & Kellaghan, 2008). Currently, national and international assessment programmes are mainly used to monitor and evaluate the quality of student learning outcomes (Postlethwaite & Kellaghan, 2008) and are designed to enable comparisons over time (Greaney & Kellaghan, 2008) although in some countries national assessments still serve a 'gate-keeping' function due to limitations in 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 (Forster 2001; Braun & Kanjee, 2006; Benavot & Tanner, 2007), although the appropriateness of using standardised tests in this way has been questioned (e.g. Goldstein & 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 & 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 civics, are principally the same curricular areas assessed in national assessments, as these subjects constitute the majority of curricula in primary education cross-nationally (Kamens & 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 more similar economic and social conditions (Kamens & Benavot, 2011) in order to explicitly compare student achievement cross-nationally. In this review, regional assessments will be understood to be a type of international assessment programme.

For this review, national and international assessment programmes will be understood to be assessment programmes conducted in primary and secondary education, and will exclude assessment programmes that are undertaken at the sub-national level (e.g. state, province, district). Furthermore, references to non-standardised assessments will not be considered in this review.

1.2.2 Education policy

The review seeks 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 understand the education policy-making processes (Haddad, 1995).

In a United Nations Educational, Scientific and Cultural Organization- International Institute for Educational Planning (UNESCO-IIEP) booklet from a series on 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.” (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, 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 idea of the policy cycle - seeing the process as having separate stages. A number of models of the policy cycle have been proposed, generally involving six to eight stages (Brigdman & Davis, 2004; Haddad, 1995; Young & Quinn, 2002). This review discusses educational policy-making by using a simplified policy cycle model from Sutcliffe and Court (2005):

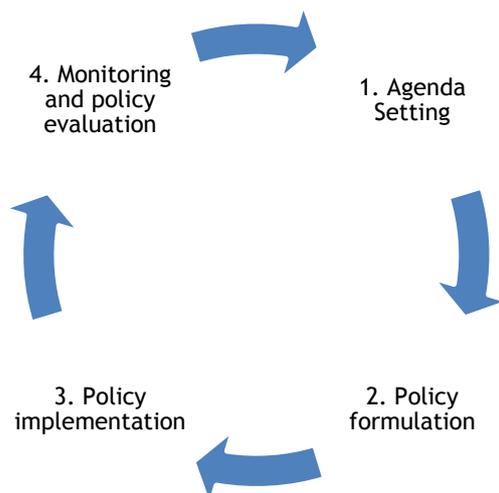


Figure 1.2.2: Simplified Model of the Policy Cycle; Source: Sutcliffe and Court (2005)

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;
- **Monitoring and Policy Evaluation:** the nature of monitoring and evaluation of policy need, design, implementation and impact.

Assessment data can be used at different stages of the 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 uses an inclusive classification in defining “developing countries”. This combines the Australian Agency for International Development’s (AusAID) 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) and World Bank’s list of low- and middle-income countries (extracted in November 2010). The World Bank classifies low- and middle income countries by using gross national income per capita. The combined list includes 157 countries, 13 of which are unique to the AusAID list and 6 are unique to the World Bank list. Please see Appendix 1.2.3 for the complete classification list of developing countries.

1.2.4 Resource allocation

Resource allocation refers to the resources that schools receive, which are frequently defined as inputs (Hanushek, 2003). These may include traditional measures of resources such as expenditure per student (Hanushek & Kimko, 2000)

or national educational budgets. Resources may also 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). Other resources may include other class and school level characteristics such as class size (Kreuger, 2002), teacher-to-student ratios and instructional time (Woessmann, 2000). Resources may also refer to teacher characteristics such as teacher experience and level of teacher qualifications (ibid.). 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 to improve school and classroom level factors such as teaching and learning strategies. Teaching and learning practices are more able to be affected by policy interventions than other factors related with student learning outcomes, such as student, family and community background characteristics (OECD, 2009a).

Teaching practices have been internationally conceptualised in the OECD's Teaching and Learning International Survey (TALIS). Using a policy framework, twenty-four 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, extra-curricular activities as well as motivation and future plans (OECD, 2009b).

1.3 Theory, policy and practice background

This section describes several 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 & Watts, 2003), there is a general understanding that

the approach involves the “rational, rigorous and systematic” (Sutcliffe & 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’ (ibid.; Sackett et al., 1996). This approach then permeated other policy sectors and fields of practice, including education, social work and criminal justice (Solesbury, 2001). This approach and the term was most prominently adopted by the UK government in the late 1990s (Sutcliffe & Court 2005).

The most common criticism to 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 (ibid., Marston & Watts, 2003; Wiseman, 2010). This review, however, does not directly engage with this issue as it is concerned with only one specific type of data - that from system-level assessments. The way that data is analysed, however, is still relevant and this forms one of the sub-questions to this review.

Other concerns with the practice of evidence-based policy-making relates 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, and less present at the stage of designing policies and programmes (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) have outlined these different issues in a table that has been reproduced as Table 1 below.

Table 1.3.1: Components of the policy cycle and evidence issues

Source: Sutcliffe and Court (2005, adapted Pollard & Court, 2005)

Stage of the policy cycle	Description	Evidence issues
Agenda setting	Awareness and priority given to an issue	The evidence needs here are in terms of 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. A key factor here is the credibility of evidence but 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: 13-14).	For both stages, policymakers 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 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 is important.
Policy implementation	Actual practical activities	Here the focus is on operational evidence to improve the effectiveness of initiatives. This can include analytic 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 here 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.

The above table serves as a synthesising framework in this review to examine the policy cycle stages in which assessment data are utilised as evidence. This framework will further examine the how the ways in which assessment data 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 & 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 (ibid.). These are increasingly being overcome, leading to a greater focus on evidence-based policy processes in developing countries (ibid.). This is one possible explanation to 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 (UNESCO 2008; Benavot & Tanner, 2007), as well as international and regional learning assessments (Kamens & McNeely, 2009). Much of this growth, especially in national learning assessments, has occurred in economically developing countries (Postlethwaite & Kellaghan, 2008).

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

There has been a shift from 1960 to 2008 from the majority of developing countries participating in international assessments to national assessments, though there has been growth in the number of assessments conducted in all three types of assessment programmes (ibid.). Therefore by 2008, national assessments constitute slightly under two-thirds of all assessment programmes undertaken by developing countries. Descriptive results from this study suggest possible regional differentiation in country participation by these 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 Sub-Saharan Africa. Examining regional assessments by regional participation, this relationship reverses and developing countries in Sub-Saharan Africa demonstrate the highest participation rates, followed by developing countries in Latin America and the Caribbean. Furthermore, considering participation in national assessments, countries in Latin America and the Caribbean have the highest participation rates, and all other regions have participation rates about 50%, except for developing countries in Eastern Europe and Central Asia.

This growth 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 that the development of human capital, as measured by learning assessments, is related to a country's economic growth (Hanushek & Kimko, 2000). Assuming a neo-institutional perspective, the rapid uptake of these assessment programmes in developing countries may be explained.

There exists a world culture among a world-polity of countries (Boli & Thomas, 1997). Actors, such as nongovernmental organisations (NGOs) and international nongovernmental organisations (INGOs), disseminate global values and subsequent

norms, which countries enact to gain legitimacy in the world-polity (Meyer, Boli, Thomas & Ramirez, 1997). The enactment of global norms is especially important for developing countries, as they are accountable to adhere to global values (Benavot & Kamens, 2011). In the world polity, there has been a dominance of the human rights discourse (Tsutsui & Wotipka, 2004), which also extends to educational institutions (Suarez, 2007), and the human rights imperative of quality education for all. The use of assessments to monitor and evaluate quality of education can be seen as an extension of the dominance and institutionalisation of a scientific discourse, in which scientific instruments, such as assessments, have a global cultural authority (Drori, Meyer, Ramirez & Schofer, 2005). Therefore these theorised mechanisms propose that INGOs and NGOs aid to diffuse the global value of human rights and quality education, and developing countries are held accountable in the world polity to use scientific instruments to monitor their adherence to these norms. There is a descriptive portrait of developing country participation in assessment programmes, and theorised mechanisms to explain the rapid cross-national growth in the use of these assessments. Little is known about the effects of such assessment programmes on education policy and practices in developing countries, still a question that this review seeks to address.

1.3.3 Assessment data as evidence in education policy planning

Evidence-based educational 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 has provided support, among others, to an emphasis on the use of student assessment data in the policy process (Campbell & 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 the outcomes of 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 (Forster 2001; Braun & Kanjee, 2006; Benavot & Tanner, 2007; Postlethwaite & Kellaghan, 2008).

Data resulting from assessment programmes can be reported to describe the extent to which an education system is teaching its students what is expected, differences in these achievement levels by sub-groups (such as gender or region) and, if background data are collected, factors that contribute to reaching different levels of achievement. Either as part of the official reporting process or resulting from secondary analysis (e.g. Lietz, Wagemaker, Neuschmidt & Hencke, 2008), this information is often also utilised to formulate recommendations to improve education outcomes.

In discussing the uses of evidence for educational policy making, Wiseman (2010) outlined three main goals: evidence-based policy making for measuring and ensuring quality, for ensuring equity, and for control, which will be referred to throughout the 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 upon 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 anticipate finding examples of the uses of assessment data as evidence in policy making in developing countries to fall into the following main groups:

- Assessment outcomes as a measure of quality level, to diagnose the strengths and weaknesses of a system. The analysis that this use relies on will likely be achievement level analysis, comparison between sub-sections in the assessment content, and trend analysis noting changes in achievement level over time. This type of use will likely be present during the agenda setting and policy formulation stages of the policy cycle, and 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.
- Assessment data used to measure and ensure equity within the system. This is likely to rely on analysis that provides comparisons between groups (schools, regions, socio-economic groups) in the system. This use will 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.
- The use of assessment data as an accountability tool, or assessment as evidence to practice 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 stakeholder, including the government, as well as policy makers using changes in assessment results to monitor the outcome of their interventions. We anticipate 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 likely take place during the monitoring and evaluation stage of the policy process.

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 formulation of a policy, assessment outcomes may demonstrate difficulties in a certain intervention, rather than feasibility (Husen & 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 ripe 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 (Hardy & Boyle, 2011; Bantick, 2011;

Topsfield, 2011) and, in the United States, 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, Greaney & Murray, 2009). As little is also known even on how assessments are used in policy making in developing countries (ibid.), 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 seeks 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 possible such 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 & Kanjee, 2006; Kellaghan, Greaney & Murray, 2009)
- How well the assessment programme is integrated into existing structures, policy and decision-making processes (Kellaghan, Greaney & Murray, 2009)
- The level of involvement of policy makers in the design and implementation of the assessment programme (ibid.)

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)
- Whether analysis is undertaken with a focus on diagnosing issues in the education system, including identifying factors associated with high and low achievement (Kellaghan, Greaney & Murray, 2009)
- Factors that relate to the dissemination of findings and analysis from the assessment programme include:
 - Timeliness of results dissemination (Kellaghan, Greaney & Murray, 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 (Postlethwaite 1984; Kellaghan, Greaney & Murray, 2009)

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

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

- The effectiveness of how the education system functions (Braun & Kanjee, 2006)
- The strength of teachers' unions and their role in policy making (Kellaghan, Greaney & Murray, 2009)
- Whether there are good communication channels or distribution system 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, Greaney & Murray, 2009)
- The level of decentralisation and openness of the political system (Kellaghan, Greaney & Murray 2009; Sutcliffe & Court, 2005)
- Level of public representation and strengths of structures for aggregating and arbitrating interests in society (Kellaghan, Greaney & Murray, 2009; Sutcliffe & Court, 2005)
- Extent of academic and media freedom and the strengths of civil society (Sutcliffe & Court, 2005)
- Existence of conflict or political volatility (Sutcliffe & Court, 2005)
- Strength of accountability systems (Sutcliffe & Court, 2005)
- The role of external (multilateral and bilateral) agencies in the system (Kellaghan, Greaney & Murray, 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, have come under scrutiny in more 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, Greaney & Murray, 2009). In making this conclusion, the authors noted that limited information is available on this topic (*ibid.*).

Reviews, however, have been conducted on the impact of the International Association for the Evaluation of Educational Achievement's (IEA) Trends in Mathematics and Science Study (TIMSS) on policies in developing countries, some done 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 that national level than the local level. The evaluation report identified that policy-makers are the most important stakeholders in PISA participation and results, rather than other stakeholders such as local officials and school principals (OECD, 2009c), which could be a related factor in the smaller impact of assessment results 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 is 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 is funded by AusAID. It received one of thirty-two 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 partnership with partner governments and other development agencies, AusAID has supported the development and improvement of systems to 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 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 is being 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 brings together expertise in international assessments, in working with policy makers in developing countries, as well as in undertaking literature searches, information retrieval, and undertaking reviews and syntheses. Additionally, the team is supplemented by a systematic review expert from the Joanna Briggs Institute (JBI) at the University of Adelaide and receives support from the EPPI-Centre at the University of London. Please see Appendix 1.5.2 for detailed information of the research team.

1.5.3 Peer review and Advisory group

A Peer Review Group is hosted by the EPPI-Centre, with whom this review is registered. After providing feedback on the title registration, EPPI-Centre will also coordinate the review of this Protocol document and the draft report from this review.

A Peer Review Group will include relevant researchers, and policy-makers nominated from the organisations funding, advising and undertaking the review. These peer reviewers will have relevant systematic review, policy and topic-area interest expertise.

Peer Review Group members will include:

Peer Review Expertise	Peer Review Member	Affiliation
Systematic Review	To be nominated by EPPI-Centre	
Policy Area	To be organised by Petra Kavunenko and Mary Fearnley-Sander	Education Program Officer, AusAID; Education Program Officer, AusAID
Topic Area Interest	David Rutkowski, Ph.D	Assistant Research Scientist, Center for Evaluation and Education Policy, Indiana University

An assembled Advisory Group will provide further feedback and guidance regarding the development of the review, which may include the conceptual framework, identification of relevant literature, inclusion and exclusion criteria, and review and synthesis of the evidence. This group will provide specific feedback relevant to their expertise, which may include pertinent theoretical, policy, and practitioner knowledge of assessment programmes in developing countries. Furthermore, the potential advisory group members will be able to provide a critical voice and

knowledge of stakeholders from developing countries to better inform the relevance of this review.

Potential Advisory Group Members may include:

Name	Affiliation
Lucrecia Santibanez, Ph.D.	Economist, Center for Latin American Social Policy, RAND Corporation
Maurice Robson, Ph.D.	Chief of Education Section, Pakistan, United Nations Children's Fund
Njora Hungi, Ph.D.	The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ)

1.6 Review questions

As outlined in earlier sections, this review is aimed at examining 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 from 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, from feedback on our review proposal and title registration, reference to resource allocation and teaching and learning practices has been 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.

Along with the main question to be addressed by this review:

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

It will also seek to address the following sub-questions:

1. How are data/findings from assessment programmes being analysed in developing countries?
2. How are results from analysis of assessment data/findings being used to inform policymaking in developing countries?
3. At what stages of the policy cycle is information from assessment programmes being used in developing countries?
4. What educational policy developments/changes in developing countries have resulted from the use of assessment data/findings?
5. What are the barriers and facilitators to the use of assessment data/findings in policy making in developing countries?

It is anticipated that there will not be a significant body of literature that directly address these questions. Therefore, this review will utilise approaches that allow us to accommodate a broad range of literature and from it synthesise aspects that are relevant to answering the review questions.

1.6.1 Framework Synthesis

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 is known about their impact, and how they are being used by policymakers and practitioners. In 2009, a report published by the World Bank (Kellaghan, Greaney & Murray, 2009) found that available evidence indicated that findings from assessment programmes are not widely used, with the authors noting the limited amount of information that exist on this topic.

Our own initial exploration of the literature on this topic led to a similar early conclusion. We anticipate that the literature on this topic and in this setting will be overwhelmingly qualitative in nature. Further, we anticipate that narrative and other textual papers will form a significant part of the literature. Taking this anticipated nature of the literature and the understanding that this topic is in general also still under-researched, the use of an iterative process like the framework synthesis approach seems most appropriate to accommodate varied literature.

This approach utilises an *a priori* conceptual framework that guides the extraction and synthesis of findings (Barnett-Page & Thomas, 2009). The use of an *a priori* framework may take into account existing research, policy issues (Carroll, Booth & Cooper, 2011) as well as prior experience and knowledge (Oliver et al., 2008). Therefore the use of a deductive process, which uses an *a priori* framework, is an optimal approach to address applied policy questions (Dixon-Woods, 2011). Furthermore, this approach fulfils a pragmatic imperative by enabling researchers to search and synthesise the evidence from a large volume of literature to address timely policy issues (ibid.). As there is a body of literature regarding the types of country participation in assessment programmes, and a theoretical body of literature regarding educational policy cycles and the use of evidence-based policy-making, this approach can use 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 educational policy making and practices in developing countries.

The research team has developed an initial conceptual framework (See Figure 1.6.2 below) which incorporates the simplified model of the policy cycle into a larger framework of assessment programme impact on educational policy-making. This initial conceptual framework will be used to identify, appraise and analyse relevant literature as well as to code and synthesise them (Oliver et al., 2008).

This framework synthesis approach shifts to an iterative process in the later stages as *de novo* topics emerge from the data (Dixon-Woods, 2011). These *de novo* themes may then be incorporated into the pre-existing conceptual framework, or

they may challenge the initial conceptual framework and its initial assumed themes and topics. The analysis of the data will then be an iterative process as *de novo* themes emerge and the initial conceptual framework is modified. Therefore, the use of an initial conceptual framework, and the later development and incorporation of new themes and topics into this framework will shift the framework synthesis approach from a deductive to a more inductive analysis of the data.

1.6.2 Focus of Review

The population of interest for this review are education policymakers and practitioners in developing countries. The review seeks to understand the impact of assessment programmes on the policy-making process and practices. The actions of these groups are what lead to changes in education policies and practices in their countries. This review, therefore, concerns itself in how assessment programmes influence these actors.

Figure 1.6.2 visually depicts the *a priori* conceptual framework that was developed for this review.

The intervention that this review is concerned with are 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 the ways in which the data are analysed may impact upon how assessment findings are used by policy-makers. Concerning assessment programmes, the review will consider sub-questions one and two:

1. How are data/findings from assessment programmes being analysed in developing countries?
2. How are results from analysis of assessment data/findings being used to inform policymaking in developing countries?

Although the review is not designed in the outset to make comparisons between different types of assessment programmes, some potential comparisons that may arise from the findings include between population and sample-based assessment programmes, and between high-stakes and low-stakes assessment programmes.

The conceptual framework outlines the context, or goals of evidence-based policy-making: quality, equity and accountability, as described in section 1.3.3. What the review examines as its outcome of interest are changes in the educational policy making process (including at the policy development, implementation and evaluation stages) and in the teaching and learning practices in developing countries. Therefore the conceptual framework will consider the goals of evidence-based policy making and the stages of the policy process and associated issues in the use of evidence (see Table 1.3.1) in examining changes in policy-making processes, and particularly policies regarding teaching and learning practices. Concerning Policymakers and the Policy Process, the review will consider sub-questions three and four:

3. At what stages of the policy cycle is information from assessment programmes being used in developing countries?
4. What educational policy developments/changes in developing countries have resulted from the use of assessment data/findings?

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

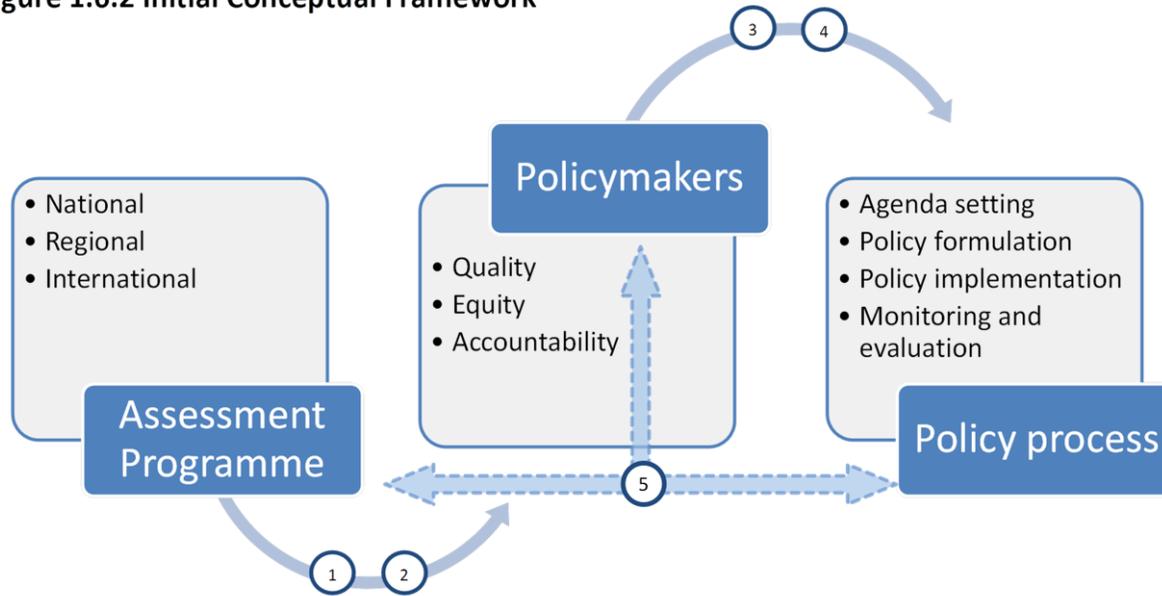
5. What are the barriers and facilitators to the use of assessment data/findings in policy making in developing countries?

on the facilitators and barriers to the use of assessment data to inform educational policy making. Possible factors may relate to the assessment programme and the analysis of results, the policy-making context and the stages in the policy cycle. These possible related factors which will be 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 are likely to address this topic and can be captured under this approach include policy analysis papers, reports on the evaluation of particular assessment programmes that describe the uses of assessment outputs and/or the impact of the assessment process, as well as textual papers, incorporating narrative and expert opinions.

Figure 1.6.2 Initial Conceptual Framework



1. How are data from assessment programmes being analysed in developing countries?
2. How are assessment findings being used by policy makers in developing countries?
3. At what stage(s) of the policy cycle is information from assessment programs being used in developing countries?
4. What educational policy developments/changes in developing countries have resulted from the use of assessment data/findings?
5. What are the barriers and facilitators to the use of assessment data/findings in policy making in developing countries?

CHAPTER TWO

2. Methods used in the review

Outline of Chapter

This chapter of the review protocol documents outlines the methodologies that will be employed for the review. The first section describes the main and anticipated users of this review and how they are going to be involved in the review processes. This is followed by the strategies for searching, including (and excluding) and reviewing the literature that will be used. A finalised version of this section will be included in the final report of this review, which will contain all necessary details to allow replication of the methods.

2.1 User involvement

The primary user 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 that are working with partner governments who are considering developing or amending or becoming involved in assessment programmes.

AusAID has 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 this review. They will also be consulted during review stages, coordinated by EPPI-Centre.

It is anticipated that this review will also 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 will be of interest to academics, researchers, outside organisations, 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.

These anticipated users are also potential sources of materials for inclusion in this study. We will therefore be involving these groups from an early stage of the review by contacting them to inform them of the review and asking for leads to potentially relevant literature. They will be identified through the authors' extensive network of stakeholders in assessment and education policy making in developing countries, and will include:

- 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
- Authors' contacts of academics in developing countries

These identified groups will also be included in the distribution list for the draft report emanating from the review, and included in the final dissemination of findings.

2.2 Identifying and describing studies

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

2.2.1 Defining relevant studies: Inclusion and exclusion criteria

Based on an initial exploration of the literature, it is anticipated that there will not be a significant body of literature that directly addresses these questions. We also consider that the methods for answering the review questions are not limited to particular study designs. Therefore, we will not be excluding studies based on pre-determined study design conditions.

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

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

Included studies will be categorised as research (including surveys, case studies, interview-based and other qualitative or quantitative papers) and non-research/textual papers (including narrative, opinion or expert papers). Based on this grouping, a second-stage analysis will be undertaken using the relevant critical appraisal checklist from JBI. The JBI set of tools were selected because it includes an appraisal tool for textual papers, which is anticipated to make up a significant portion of the literature on this topic.

This quality assessment will be used to describe the state of available evidence in this area, which is useful for recommending future areas of research. This process also allows for some analysis into the extent that studies of different types and

quality impacted on the findings of this review. It is important to reiterate, however, that this quality assessment will not be used to exclude studies.

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

As outlined in section 1.2.1, assessments may still serve a 'gate-keeping' function due to limitations in the availability of places in subsequent stages of schooling in some countries, however they are mainly used to monitor and evaluate the quality of student learning outcomes (Postlethwaite & Kellaghan, 2008) and are designed to enable comparisons over time (Greaney & 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 & Kellaghan, 2008), which only began in the 1990's.

Therefore, a report, study or paper must meet the following criteria:

- It must make reference to an assessment programme conducted in either primary or secondary education or both
- It must have been published or released between 1990 - 2011

Lastly, as outlined in section 2.2.2, searching only for English language literature will increase the possibility for publication and positive biases. We will therefore be seeking to expand the searches to non English language studies. The research teams include working knowledge in German, Spanish, French and Indonesian. Additionally, the team will seek the support of bilingual ACER colleagues to translate search terms and emails to Arabic and Chinese. Reports, studies or papers which are outside the research team's and ACER colleagues' language capabilities will be excluded from the review.

2.2.2 Identification of potential studies: Search strategy

Based on our knowledge of the literature on this topic, our search strategy will seek to identify published as well as 'grey' literature. In identifying databases for searching, we purposely include those that include grey literature. Additionally, we will also supplement our searches of databases, journals and the web with sourcing literature by contacting the groups identified in Section 2.1 above.

More specifically, we will employ five strategies for identifying potential studies:

1. Electronic searches of bibliographic databases:

Researchers will utilise databases available to ACER, including Australian Education Index, Education Resources Information Center (ERIC), Education Research Complete (ERC), British Education Index, Scopus, Eldis, Asia-Studies Full Text Online Google Scholar and British Library for Development Studies. Experienced information librarians will conduct initial searches of these sources and set up alerts to capture material indexed for the duration of the review.

The key words we will employ are firstly terms 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).

We will also employ terms for controlled-vocabulary searches using database specific key words: (“Student evaluation” “Student assessment” OR “Educational testing” OR “Achievement tests” OR “Standardized tests” OR “Standardised tests” OR “Testing programs” OR “Testing programmes” “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 will be combined with free-text terms to describe the intervention: (National exam*).

These will also be combined with controlled-vocabulary terms to narrow down the specified level of education in which the intervention is 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. The research team initially outlined to include search terms to describe types of impacts of assessment programs: (“impact” OR “effect” OR “use” OR “utilisation” OR “utilization” OR “benefit” OR “consequence”). For databases that are not focused on development or developing countries, we will supplement these with keywords that focus the search further, such as: (“developing” OR “third world” OR “impoverished”) AND (“nation” OR “country” OR “region”).

However, using terms to describe types of impacts was too restrictive, and lead to possible exclusions of relevant records for the review. Therefore, the research team decided to exclude these terms from electronic database searching, in order to increase the number of returned, relevant records. In doing so, the team recognises that many irrelevant search records may be returned, but that these papers will be excluded as the inclusion criteria are applied.

In addition, we will use the names of specific developing nations as identified by AusAID, the IMF and the World Bank, such as Indonesia, Mexico, etc., and the names of regions, such as Africa, Latin America, and Asia. If there is a geographic descriptor field in the record we will search

that field for the names of countries or regions; otherwise we will search for the names in the abstract field of the record. For databases that are not focused on education, we will add the search term “education”.

The above search strategies and search terms will need to be modified to suit each database. The appendices to our final review report will carefully document all descriptors and keywords used for each database to permit replication.

2. *Targeted searches of contents pages of key journals as new issues are published to overcome the delay between publication date and when they appear in the bibliographic databases:*

These include Assessment in Education: Principles, Policy & Practice; Educational Research; Evaluation and Research in Education; Asia Pacific Journal of Education; and International Journal of Educational Development. We will also seek to identify key regional and national journals on assessment and education policy.

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

This will include targeting known international, regional and national assessments in the repositories of agencies that manage international and regional assessment programs, such as the IEA, OECD, SACMEQ and LLECE, as well as agencies that provide support and research into assessment programs in developing countries, such as the World Bank, DFID, AusAID and UNESCO. Additionally, this also includes the publications of relevant research bodies, such as RTI, ETS, Cambridge Assessment, CREATE, ADEAnet, Campbell Collaboration, and 3ie and research bodies of key national ministries.

4. *Citation chasing:*

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

5. *Contacting relevant groups and researchers in this area:*

The groups that are identified in Section 2.1 will send query emails for potentially relevant literature. This is an iterative strategy, and key authors that have been identified through the literature search may also be contacted. ‘Snowballing’ technique will also be employed by asking contacts to refer other researchers or authors they are aware of who may have access to additional literature.

Bearing in mind the population of interest for this review, we acknowledge that limiting the searches to the English language will increase the possibility for publication and positive biases. We will therefore be seeking to expand the searches to non-English studies, using such databases and search engines as scopus

and Google Scholar. The research teams include working knowledge in German, Spanish, French and Indonesian. Additionally, the team will seek the support of bilingual ACER colleagues to translate search terms and emails to Arabic and Chinese. We believe that this will cover the major languages of publication in this setting. These searches may be further undertaken in non-English, regional databases, such as Latin American Journals Online (LAMJOL), and regional databases such as African Journals Online (AJOL), which houses peer-reviewed articles from Southern scholars.

A database system will be set up to keep track of and, to code, studies found during the review. Titles and abstracts will be imported directly from the databases or and entered manually, when necessary, into the first of these databases.

2.2.3 Examples of studies to be included in this review

Please see Appendix 2.2.3 for an applied example of the inclusion and exclusion criteria.

Crespo, Soares & de Mello e Souza (2000): This study analyses the impact of the Brazilian National System of Evaluation of Basic Education on educational policies and practices, especially at the state level. It utilises data from interviews, site visits, document review, and a survey completed by state education departments and identifies the strong points and weaknesses of the evaluation system.

Gilmore (2005): This study examines the impact of participation in the Progress in International Reading Literacy Study (PIRLS) and TIMSS on low- and middle-income countries, serving as an evaluation of the value of World Bank support to these countries. Data is obtained by the author herself, including through observing a four-day meeting for each of PIRLS and TIMSS; interviews with the National Research Coordinator (NRC) of each World Bank-funded country and questionnaires sent to the NRC and a nominated senior education official of each country.

Leste (2005): This paper is a presentation by an official from the Ministry of Education of the Republic of Seychelles at a SACMEQ Research Conference. It describes the way SACMEQ data was 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.

Nzomo & Makuwa (2006): This book chapter described the processes undertaken by the Ministries of Education in Kenya and Namibia in utilising SACMEQ findings to, among others, undertake modifications to the curriculum, allocate budget to monitor education quality and develop programmes to improve greater efficiency in the education system. The authors are current and former National Research Coordinators in the two countries and extensively cite government policy documents.

2.2.4 Examples of studies to be excluded from this review

Braun & Kanjee (2006): This paper provides a framework of how national assessments can impact upon an education system and discusses the potential uses and impact of national assessments on education policy in developing countries.

However, it does not make any reference to actual or existing link between national assessment findings and policy changes.

Lubisi & Murphy (2002): This article reviews assessment policy and practice in South African schools in recent years and overview of historical changes. The focus is on classroom-based assessments and not standardised assessments; therefore it will be excluded from this review.

Sunderman & Orfield (2008): This study examines the response of state education departments, that are tasked with designing interventions in underperforming schools in the mandated yearly standardised tests in elementary and middle school under the No Child Left Behind Act. The authors collect data from interviews, analysis of policy and programme documents, and budget and staffing information. As it only references the United States, however, it will be excluded from this review.

2.2.5 Screening studies: Applying inclusion and exclusion criteria

Two research assistants and an ACER librarian will undertake the database searches using the strategies outlined above. Using formalised search strategies, all identified titles, and when available, abstracts, will be screened for relevance to the review criteria. Where a citation is not accompanied by an abstract, if the title suggests relevance, the full paper will be retrieved and analysed for relevance to the review criteria.

To re-iterate, the application of this process to the literature will be based on the goodness of fit with the review criteria rather than the methodological quality of the studies. Following the database specific analysis of titles and available abstracts, literature considered relevant to the review criteria will be downloaded in full and stored in the EPPI Reviewer 4 software. Each paper will subsequently be subject to further screening for relevance to the review criteria according to the methods following below.

2.2.6 Characterising included studies (if EPPI-Centre review: EPPI-Centre and review-specific keywording)

Following initial screening of titles and abstracts and the application of inclusion and exclusion criteria, remaining papers will be descriptively mapped, or keyworded according to the using EPPI-Centre (2003) Core Keywording Strategy: version 0.9.7, which is modified to include additional keywords which are specific to the context of the review. These additional keywords will identify the type of assessment programme referred to in the study and the type of impact that is described. The type of impact will be keyworded by identifying the stage of the policy process influenced by assessment, the goals in the use of the assessment, and the facilitators or barriers in the use of the assessment.

The preliminary keywording strategy to be used for this review is presented in Appendix 2.2.6. This keywording strategy may be modified upon identification of included studies. The final keywords used will be listed in the final report of this review.

2.2.7 Quality assurance process for screening and keywording studies

This Review group will aim to meet quality assurance of the outlined processes with the EPPI-Reviewer 4 software. This software will enable the Review Group to import searches, apply inclusion and exclusion codes, and apply keywords to the imported studies. The EPPI-Reviewer 4 software will enable our review group to ensure quality assurance by maintaining transparent, documented processes and be able to negotiate a shared and applied understanding of the methods and concepts under review by all team members.

2.3 In-depth review

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

After the inclusion and exclusion criteria have been applied to the data, and keyworded, the Review Group will apply the initial conceptual framework to the data. While the application of an *a priori* framework will analyse data according to pre-determined themes, moving to an in-depth review will allow for new concepts and themes to emerge, or pre-existing themes within the framework may be challenged. This in-depth review will require an iterative process of analysis. As new themes are identified and defined, or the framework changed, the new framework must be applied to the literature. This iterative approach was previously outlined in section 1.6.1:

“This framework synthesis approach shifts to an iterative process in the later stages as *de novo* topics emerge from the data (Dixon-Woods, 2011; Carroll et al., 2011). These *de novo* themes may then be incorporated into the pre-existing conceptual framework, or they may challenge the initial conceptual framework and its initial assumed themes and topics. The analysis of the data will then be an iterative process as *de novo* themes emerge and the initial conceptual framework is modified. Therefore, the use of an initial conceptual framework, and the later development and incorporation of new themes and topics into this framework will shift the framework synthesis approach from a deductive to a more inductive analysis of the data.”

As the literature is analysed according to the framework, data will be aggregated under the themes, or concepts of this framework. During this iterative process, as the framework is modified, the data will be organised and re-aggregated under the modified themes.

As little is known regarding the impact of national and international assessment programmes, this iterative process will be an important step to enable our review group to move to an in-depth review of the literature with a better understanding of the existing literature. This revised conceptual framework will then be applied to the keyworded literature.

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

Two reviewers will independently use a standardised tool to extract data from the included studies. Detailed data from studies will be extracted on the following:

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

In addition to the above information, the reviewers will also apply a standardised quality assessment to the included studies (described in section 2.3.3 below).

2.3.3 Assessing quality of studies and evidence for the review question

Included studies will be categorised as research (including surveys, case studies, interview-based and other qualitative or quantitative papers) and non-research/textual papers (including narrative, opinion or expert papers). Based on this grouping, a second-stage analysis will be undertaken using the relevant critical appraisal checklists that are modified from those developed by JBI. The JBI set of tools were selected because it includes an appraisal tool for textual papers, which is anticipated to make up a significant portion of the literature on this topic.

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 of interpretation of data?

For non-research/textual papers:

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

This quality assessment will be used to describe the state of available evidence in this area, which is useful for recommending future areas of research. This process also allows for some analysis into the extent that studies of different types and quality impacted on the findings of this review (See Section 2.3.5).

2.3.4 Synthesis of evidence

Using the method employed by Oliver et al. (2004) in their framework synthesis as a guide, our review team may use charts to create 'typologies' of the themes included in the revised conceptual framework. This will involve the creation of summaries of all the relevant literature for that theme or concept included in the framework. These summaries, or 'typologies' will be organised in charts, or a synthesis table.

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

An example of a preliminary synthesis framework, which can be used to create typologies of themes, and then describe relationships between and across themes, is provided in Appendix 2.3.4.

2.3.5 In-depth review: Quality assurance process

We will conduct a secondary 'sensitivity-analyses' for literature that was appraised as being of high quality using the JBI Appraisal Tools. To ensure the quality and soundness of our conclusions based on the anticipated diverse body of literature, we will then conduct sensitivity analyses to assess if our results also apply to a body of high-quality literature.

CHAPTER THREE

3. References

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 P. Lietz, H. Wagemaker, O. Neuschmidt, & J. Henke (Eds.), *Educational issues in the Middle East North Africa region* (pp. 57-76). Hamburg: International Association for the Evaluation of Educational Achievement.

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

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

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. *Paper commissioned for the EFA Global Monitoring Report 2008, Education for All by 2015: will we make it?* (pp.1-17). Paris: United Nations Educational, Scientific and Cultural Organization.

Berry, R., & Adamson, B. (2011). Assessment reform past, present and future. In R. Berry & B. Adamson (Eds.), *Assessment reform in education, Education in the Asia-Pacific region: Issues, Concerns and prospects* (pp.3-14). Springer: New York.

Boli, J. & Thomas, G.M. (1997). World culture in the world polity: A century of international non-governmental organization. *American Sociological Review*, 62, 171-190.

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

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

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

Campbell, C., & Levin, B. (2009). Using data to support educational improvement. *Educational assessment, evaluation and accountability*, 21, 47-65.

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 chemopreventive agents. *BMC Medical Research Methodology*, 11 (29), 1-9.

Crespo, M., Soares, J.F. & de Mello e Souza, A. (2000). The Brazilian national evaluation system of basic education: Context, process and impact. *Studies in Educational Evaluation*, 26 (2), 105-25.

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 the 4th Annual Campbell Collaboration Colloquium, Washington DC.

deLandshere G. (1997). History of educational research. In J.P. Keeves (Ed.), *Educational research, methodology and measurement: An international handbook* (2nd ed.) (pp. 8-16). Cambridge, UK: Pergamon Press.

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

Drori, G.S., Meyer, J.W., Ramirez, F.O. & Schofer, E. (2003). *Science in the modern world polity: Institutionalization and globalization*. Stanford, CA: Stanford University Press.

Forster, M. (2001). *A policy maker's guide to systemwide assessment programs*. Camberwell: The Australian Council for Educational Research Ltd. Retrieved November 1, 2010 from Research Publications at ACEReSearch site: 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, S.M. (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: The World Bank.

Haddad, W.D. (1995). *Education policy-planning process: An applied framework*. Paris: UNESCO-International Institute for Educational Planning.

Hanushek, E.A. (2003). The failure of input-based schooling policies. *The Economic Journal*, 113, F64-F98.

Hanushek, E.A., & Kimko, D.D. (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.

Husen, T. & Kogan, M. (1984). *Educational research and policy: How do they relate?* Oxford: Pergamon Press Ltd.

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

Kamens, D.H. & McNeely, C.L. (2009). Globalization and the growth of international educational testing and national assessment. *Comparative Education Review*, 54(1), 5-25.

Keeves, J. P. (1995). *The World of School Learning: Selected Key Findings from 35 Years of IEA Research*. The Hague, Netherlands: International Association for the Evaluation of Educational Achievement.

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

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, A.B. (2002). Understanding the magnitude and effect of class size on student achievement. In L. Mishel & R. Rothstein (Eds.), *The class size debate* (pp. 7-35). Washington DC: The Economic Policy Institute.

Leste, A. (2005). Streaming in Seychelles: From SACMEQ research to policy reform. Paper presented at the 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*. IEA: Netherlands.

Lubisi, R.C. & Murphy, R.J.L. 2002. Assessment in South African Schools. *Assessment in Education: Principles, Policy & Practice*, 9 (2), 255-268.

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.

Meyer, J.W., Boli, J., Thomas, G.M. & Ramirez, F.O. (1997). World society and the nation-state. *American Journal of Sociology*, 103(1), 144-81.

Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., Arora, A., & Erberber, E. (2005). *TIMSS 2007 assessment frameworks*. Chesnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Nzomo, J. & Makuwa, D. (2006). Chapter 10: How can countries move from dissemination, and then to policy reform? (Case studies from Kenya and Namibia).

In K. Ross & I.J. Genevois (Eds.), *Cross-national studies of the quality of education* (pp. 213-228). Paris: UNESCO-IIEP.

Oliver, S. R., Rees, R. W., Clarke-Jones, L., Milne, R., Oakley, A. R., 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.

Organisation for Economic Co-Operation and Development (2009a). *Effective teaching and learning environments: First results from TALIS*. Paris: OECD Publishing.

Organisation for Economic Co-Operation and Development (2009b). *PISA 2009 assessment framework: Key competencies in reading, mathematics and science*. Paris: OECD Publishing.

Organisation for Economic Co-Operation and Development (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.

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, W.J. (1999). Why standardized test scores don't measure educational quality. *Educational Leadership*, 56(6), 8-15.

Postlethwaite, T.N. & Kellaghan, T. (2008). *National assessments of educational achievement. Education policy series 9*. Paris: The International Institute for Educational Planning (IIEP) & Brussels: The International Academy of Education (IAE).

Postlethwaite, T.N. (1984). Research and policy-making in education: Some possible links. In T. Husén & M. Kogan (Eds.), *Educational research and policy: How do they relate?* (pp. 195-206). Great Britain: Pergamon Press.

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

Sackett, D.L., Rosenberg W.M., Gray JA, Haynes R.B. & Richardson W.S. (1996). Evidence based medicine: What it is and what it isn't [Editorial]. *BMJ*, 312, 71-2.

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.

Suarez, D. (2007). Education professionals and the construction of human rights education. *Comparative Education Review*, 51(1), 48-70.

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

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, March 12). Labor may have won education wars, but it is a pyrrhic victory. *The Age*. Retrieved online August 5, 2011, from <http://www.theage.com.au/opinion/politics/labor-may-have-won-education-wars-but-its-a-pyrrhic-victory-20110311-1br8q.html>

Tsutsui, K. & Wotipka, C.M. (2004). Global civil society and the international human rights movement: Citizen participation in human rights international nongovernmental organizations. *Social Forces*, 83 (2), 587-620.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2008). *Education for all global monitoring report. Education for All by 2015: Will we make it?* Paris: UNESCO.

Wiseman, A.W. (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.

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.

APPENDIX 1.1: Authorship of this report

Authors (in alphabetical order):

Ms. Pat Knight, ACER

Dr. Petra Lietz, ACER

Ms. Dita Nugroho, ACER

Ms. Mollie Tobin, ACER

Details of Peer Review Group membership:

Peer Review Expertise	Peer Review Member	Affiliation
Systematic Review	To be nominated by EPPI-Centre	
Policy Area	To be organised by Petra Kavunenko and Mary Fearnley-Sander	Education Program Officer, AusAID; Education Program Officer, AusAID
Topic Area Interest	David Rutkowski, Ph.D	Assistant Research Scientist, Center for Evaluation and Education Policy, Indiana University

Details of Potential Advisory Group membership:

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

Acknowledgements:

We wish to acknowledge the feedback provided by Craig Lockwood of the Joanna Briggs Institute (JBI) on a draft of this protocol and to thank him for his helpful contributions regarding this methodology chapter and following sections. His feedback was directly incorporated in section 2.2.5.

Conflicts of interest:

The authors have no financial interests to disclose. We have some experience in the conduct of national assessment studies in Indonesia (Kos, Nugroho & Lietz, 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 on 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.

APPENDIX 1.2.3: Classification of developing countries

Table 1.2.3 : Inclusive Classification of Developing Countries (N= 157)

AusAID ¹	World Bank ²
Afghanistan	Afghanistan
Albania	Albania
Algeria	Algeria
	American Samoa
Angola	Angola
Anguilla	
Antigua and Barbuda	Antigua and Barbuda
Argentina	Argentina
Armenia	Armenia
Azerbaijan	Azerbaijan
Bangladesh	Bangladesh
Barbados	
Belarus	Belarus
Belize	Belize
Benin	Benin
Bhutan	Bhutan
Bolivia	Bolivia
Bosnia and Herzegovina	Bosnia and Herzegovina
Botswana	Botswana
Brazil	Brazil
	Bulgaria
Burkina Faso	Burkina Faso
Burma (Myanmar)	Burma (listed as Myanmar)
Burundi	Burundi
Cambodia	Cambodia
Cameroon	Cameroon
Cape Verde	Cape Verde
Central African Republic	Central African Republic
Chad	Chad
Chile	Chile

(Continued)

Table 1.2.3 (Continued).

AusAID ¹	World Bank ²
China, (excluding Hong Kong)	China
Colombia	Colombia
Comoros	Comoros
Congo, Democratic Republic	Congo, Democratic Republic
Congo, Republic	Congo, Republic
Cook Islands	
Costa Rica	Costa Rica
Côte d'Ivoire	Côte d'Ivoire
Croatia	
Cuba	Cuba
Djibouti	Djibouti
Dominica	Dominica
Dominican Republic	Dominican Republic
Ecuador	Ecuador
Egypt, Arab Republic	Egypt, Arab Republic
El Salvador	El Salvador
Equatorial Guinea	
Eritrea	Eritrea
Ethiopia	Ethiopia
Fiji	Fiji
Gabon	Gabon
Gambia, The	Gambia, The
Georgia	Georgia
Ghana	Ghana
Grenada	Grenada
Guatemala	Guatemala
Guinea	Guinea
Guinea-Bissau	Guinea-Bisau
Guyana	Guyana
Haiti	Haiti
Honduras	Honduras
India	India
Indonesia	Indonesia
Iran, Islamic Republic	Iran, Islamic Republic
Iraq	Iraq
Jamaica	Jamaica
Jordan	Jordan
Kazakhstan	Kazakhstan
Kenya	Kenya
Kiribati	Kiribati
Korea, Democratic Republic	Korea, Democratic Republic

(Continued)

Table 1.2.3 (Continued).

AusAID ¹	World Bank ²
	Kosovo
Kyrgyz Republic	Kyrgyz Republic
Laos, People's Democratic Republic	Laos, People's Democratic Republic
Lebanon	Lebanon
Lesotho	Lesotho
Liberia	Liberia
Libya	Libya
	Lithuania
Macedonia, Former Yugoslav Republic	Macedonia, Former Yugoslav Republic
Madagascar	Madagascar
Malawi	Malawi
Malaysia	Malaysia
Maldives	Maldives
Mali	Mali
Marshall Islands	Marshall Islands
Mauritania	Mauritania
Mauritius	Mauritius
Mayotte	Mayotte
Mexico	Mexico
Micronesia, Federated States of	Micronesia, Federated States of
Moldova	Moldova
Mongolia	Mongolia
Montserrat	
(Jointly listed with Serbia)	Montenegro
Morocco	Morocco
Mozambique	Mozambique
Namibia	Namibia
Nauru	
Nepal	Nepal
Nicaragua	Nicaragua
Niger	Niger
Nigeria	Nigeria
Niue	
Oman	
Pakistan	Pakistan
Palau	Palau
Panama	Panama
Papua New Guinea	Papua New Guinea
Paraguay	Paraguay
Peru	Peru
Philippines	Philippines
	Romania
	Russian Federation

(Continued)

Table 1.2.3 (Continued).

AusAID ¹	World Bank ²
Rwanda	Rwanda
Samoa	Samoa
São Tomé and Príncipe	São Tomé and Príncipe
Senegal	Senegal
Serbia and Montenegro	Serbia
Seychelles	Seychelles
Sierra Leone	Sierra Leone
Solomon Islands	Solomon Islands
Somalia	Somalia
South Africa	South Africa
Sri Lanka	Sri Lanka
St. Helena	
St. Kitts and Nevis	St. Kitts and Nevis
St. Lucia	St. Lucia
St. Vincent and the Grenadines	St. Vincent and the Grenadines
Sudan	Sudan
Suriname	Suriname
Swaziland	Swaziland
Syrian Arab Republic	Syrian Arab Republic
Tajikistan	Tajikistan
Tanzania	Tanzania
Palestine (Territories Administered by the	West Bank and Gaza
Thailand	Thailand
Timor-Leste (East Timor)	Timor-Leste
Togo	Togo
Tokelau	
Tonga	Tonga
Trinidad and Tobago	
Tunisia	Tunisia
Turkey	Turkey
Turkmenistan	Turkmenistan
Tuvalu	Tuvalu
Uganda	Uganda
Ukraine	Ukraine
Uruguay	Uruguay
Uzbekistan	Uzbekistan
Vanuatu	Vanuatu
Venezuela	Venezuela, RB

(Continued)

Table 1.2.3 (Continued).

AusAID ¹	World Bank ²
Vietnam	Vietnam
Wallis & Futuna	
Yemen, Republic	Yemen, Republic
Zambia	Zambia
Zimbabwe	Zimbabwe

¹ AusAid. 2009. *List of Developing Countries: As Declared By The Minister For Foreign Affairs.* (http://www.ausaid.gov.au/ngos/devel_list.cfm).

² World Bank. 2010. *Country and Lending Groups.* (<http://data.worldbank.org/about/country-classifications/country-and-lending-groups>).

APPENDIX 2.2.3: Inclusion and exclusion criteria

Basic Study Details	Author	Gilmore	Leste
	Year	2005	2005
	Title	THE IMPACT OF PIRLS (2001) AND TIMSS (2003) IN LOW- AND MIDDLE-INCOME COUNTRIES	Streaming in Seychelles: From SACMEQ Research to Policy Reform
	Citation	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)." International Association for the Evaluation of Educational Achievement, Amsterdam.	Leste, A. (2005)." Streaming in Seychelles: From SACMEQ Research Policy Reform". Paper presented to the International Invitational Educational Policy Research Conference, Paris, France, 28 September 2 October, 2005.
	Assessment Program Type	International	Regional
	Assessment Program Name	PIRLS and TIMSS	SAQMEC
	Location	Global	Seychelles
	Type of Publication	Evaluation report	Presentation (SAQMEC Research Conference)
	Inclusion Criteria	Reference is made to a national, regional or international standardised assessment or testing programme.	Yes
Refers to at least one developing country		Yes	Yes
An explicit reference to the link between an assessment programme and a stage in the policy making process (policy design; evaluation of policy options; policy implementation; policy evaluation); OR		Yes	Yes
a change in policy within the education sector (e.g. allocation of resources in education, curriculum design, standards development); OR		n/a	n/a
a change in teaching or learning practices at the classroom or school level		n/a	Yes
Published/released between 1990 - 2011		Yes	Yes
In English, French, German, Indonesian, Spanish, Arabic or Chinese		Yes	Yes
INCLUDE IN REVIEW?	YES	YES	

APPENDIX 2.2.6: EPPI-Centre keyword sheet including review-specific keywords

Review-specific keywords

B.1 At what level is the assessment program implemented?	B.1.1 National B.1.2 Regional B.1.3 International
B.2 What is the sampling strategy of the assessment programme?	B.2.1 Population/Census B.2.2 Representative sample B.2.3 Convenience sample
B.3 What stage of the policy process is influenced by assessment as described in the study?	B.3.1 Agenda setting B.3.2 Policy formulation B.3.3 Policy implementation B.3.4 Monitoring and policy evaluation
B.4 What is the goal of the use of assessment described in the study?	B.4.1 Measure and ensure quality B.4.2 Measure and ensure equity B.4.3 Accountability
B.5 What facilitators or barriers to the use of assessment are described in the study?	B.5.1 Nature of assessment programme B.5.2 Analysis of outcomes B.5.3 Dissemination of findings B.5.4 Nature of education system B.5.5 Wider political context

APPENDIX 2.3.4: Preliminary synthesis table

		Quality		Equity		Accountability	Goal
		Agenda setting	Policy formulation		Policy implementation	Monitoring and policy evaluation	Policy stage
National	Census	A	B	C	D	E	
	Sample	F	G	H	I	J	
Regional		K	L	M	N	O	
International		P	Q	R	S	T	

Assessment type *Sampling approach*

First produced in 2012 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 6367

<http://eppi.ioe.ac.uk/>

<http://www.ioe.ac.uk/ssru/>

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.

The EPPI-Centre was established in 1993 to address the need for a systematic approach to the organisation and review of evidence-based work on social interventions. The work and publications of the Centre 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.

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