The development of SAGE: A tool to evaluate how policymakers’ engage with and use research in health policymaking

Steve R. Makkar1,*, Sue Brennan2, Tari Turner2, Anna Williamson1, Sally Redman1 and Sally Green2

1The Sax Institute, Level 13, Building 10, 235 Jones Street Ultimo NSW 2007 and 2School of Public Health and Preventive Medicine, Monash University, Level 6, The Alfred Centre, 99 Commercial Road Melbourne VIC 3004

*Corresponding author. Email: steve.makkar@saxinstitute.org.au

Abstract

It is essential that health policies are based on the best available evidence including that from research, to ensure their effectiveness in terms of both cost and health outcomes for the wider community. The present study describes the development of SAGE (Staff Assessment of enGagement with Evidence), a measure that combines an interview and document analysis to evaluate how policymakers engaged with research (i.e., how research was searched for, appraised, or generated, and whether interactions with researchers occurred), how policymakers used research (i.e., conceptually, instrumentally, tactically, or imposed), and what barriers impacted upon the use of research, in the development of a specific policy product. A multifaceted strategy was used to develop the SAGE interview and the accompanying interview-scoring tool. These included consultations with experts in health policy and research, review and analysis of the literature on evidence-informed policymaking and previous measures of research use, qualitative analysis of interviews with policymakers, and pilot-testing with senior policymakers. These steps led to the development of a comprehensive interview and scoring tool that captures and evaluates a broad range of key actions policymakers perform when searching for, appraising, generating, and using research to inform a specific policy product. Policy organizations can use SAGE to not only provide a thorough evaluation of their current level of research engagement and use, but to help shed light on programs to improve their research use capacity, and evaluate the success of these programs in improving the development of evidence-informed policies.

Key words: evidence-informed policy; knowledge translation; public health; measurement; evidence-based policy; health policy.

Introduction

There is growing emphasis on the importance of utilizing the best available evidence, particularly from research, to inform policy development and help ensure state and national objectives for improved health and more efficient health spending are achieved (Hanney et al. 2003; Brownson et al. 2009a,b; LaRocca et al. 2012). Indeed, research has helped to inform policies that have been associated with improvements in health in a range of areas (Buchan 2004; Hanna et al. 2004; Fielding and Briss 2006; Morrato et al. 2007; Andre et al. 2008; Bowen et al. 2009; Milat et al. 2013).

Numerous studies, however, indicate that globally there remains a considerable gap between evidence of effective strategies and the health policies that are developed and implemented. Several empirical studies of decision makers in the public health field have documented the underuse of research to guide the formulation of policies, programs, and other decisions (Trostle et al. 1999; Amara et al. 2004; Buchan 2004; Brownson et al. 2009a; Ritter 2009; Chagnon et al. 2010). Policymaking, however, is a complex process, wherein research is one of many contributing factors alongside other influences such as political factors, interests of key stakeholders,
feasibility (e.g., availability of resources), and other sources of evidence (e.g., locally collected data, past policy documents; Black 2001; Brownson et al. 2009b). It is important, therefore, that policymakers’ decisions are informed by, as opposed to being entirely based on, evidence from research.

The importance of measuring research use

In this context, it is essential that validated measures of research use in health policymaking are developed (Dobbins et al. 2002; Boaz et al. 2009). Organizations that develop health policies and programs could use such measures to assess their current level of engagement with, and use of, research. In particular, these measures could be used to evaluate the degree to which staff access and use systematic reviews, which arguably provide the most reliable, efficient, and accurate information about ‘what works’ to inform policymaking (Lavis et al. 2005; Moat et al. 2013).

The results of such measurement may help select strategies, tools, and programs to improve organizations’ research use capacity. Measures of research use could then be used to monitor and evaluate the impact of these strategies on staff research use, as well as providing information needed to explore the relationship between research use and health outcomes and expenditures (Redman et al. 2015). Evidence that research is in fact being used, and is associated with health gains and reduced spending, may increase commitment to the use of research in policy development, promote funding and production of policy-relevant research by research organizations (Hanney et al. 2003; Brownson et al. 2009a,b), and justify continued government investment in policy-relevant health research (Dobbins et al. 2002).

The present study

This article describes the conceptual development of a new tool to measure research utilization known as SAGE (Staff Assessment of Engagement with Evidence). The impetus for developing SAGE came from a broader program of work of which SAGE is part. This program of work includes a trial evaluating the effectiveness of a multifaceted program to improve the capacity of health policy agencies to use research in the development of policies and programs, entitled SPIRIT (Supporting Policy in Health with Research: an Intervention Trial, see The CIPHER Investigators 2014).

In this article, we begin by establishing the need for a new measure, describing a conceptual framework for measuring research use and briefly reviewing the suitability of existing instruments for operationalizing this framework. We then describe the qualitative analysis undertaken to develop the content of SAGE.

Conceptual framework for measuring research use: The SPIRIT Action Framework

SAGE is based on the SPIRIT Action Framework which describes the steps, barriers, facilitators, and contextual influences along the pathway to research use in policymaking (Fig. 1; Redman et al. 2015). This framework underpinned SPIRIT by informing the selection of intervention strategies and outcome measures used in the trial. More generally, however, the framework aims to guide a systematic approach for selection and testing of strategies for building capacity for research use. The framework does not assume that policymaking is a linear, predictable process, but provides a simplified schematic to capture the process through which research informs policymaking. Once the need for research to inform policy is identified, policymakers initiate a number of (A) research engagement actions such as (i) searching for and (ii) obtaining research, (iii) appraising its relevance and (iv) quality, (v) generating new research or analyses, and (vi) interacting with researchers. Once relevant research has been obtained and/or generated, it can then be (B) used to inform policymaking. This may take place in four ways: (i) research may directly influence what issues to prioritize, or what decisions should be made with regard to the identified issue(s), including decisions to reject or disinvest in existing policies (instrumental); (ii) it may provide new ideas, understanding, or concepts to clarify thinking about the policy issue without directly influencing content (conceptual); (iii) it may be used to justify or lend weight to pre-existing decisions and courses of action relating to the issue, or make a case for changes to be made to existing policies (tactical); and/or (iv) be used to meet organizational, legislative, or funding requirements to use research (imposed). The framework predicts that such research use will lead to more evidence-based policies, and ultimately better health services and outcomes.

The framework also acknowledges that there are (i) individual factors (e.g., policymakers’ skills at accessing and applying research); (ii) external policy influences (e.g., media, stakeholder interests, availability of resources), and (iii) organizational factors (e.g., tools to support research use; positive research climate) that can act as either (C) barriers or facilitators to research use in policymaking. A comprehensive approach to measuring research use should ideally incorporate each of the key domains highlighted in this framework.

Suitability of existing measures for measuring SPIRIT domains

From systematic literature searches (SCOPUS, Global Health and reference lists of obtained articles) and an existing review (Squires et al. 2011), we identified many different approaches that had been used to measure research use. Six measures assessed research use in relation to health policy specifically (Landry et al. 2001a,b, 2003; Hanney et al. 2003; Amara et al. 2004; de Goede et al. 2012; Armstrong et al. 2014; Zardo and Collie 2014). The suitability of these measures for operationalizing the SPIRIT Action Framework was examined in relation to three key features: (1) coverage of each of the SPIRIT domains (i.e., assessment of multiple types of research use and engagement actions, consideration of barriers or context), (2) conceptual basis of the measure, and (3) measurement approach (e.g., self-report scale, document analysis). The strengths and limitations of previous measures with regard to these key features is described in Supplementary File 1 and summarized below.

First, with regard to coverage of the SPIRIT domains, none of the existing measures addresses policymakers’ efforts to generate or appraise research to inform policy. Assessment of research use is also limited because measures evaluate research use too broadly or focus on only one type of research use (Landry et al. 2001a,b, 2003; Armstrong et al. 2014; Zardo and Collie 2014). Some measures assess multiple types of research use but neglect imposed use (Hanney et al. 2003; Amara et al. 2004; de Goede et al. 2012). Second, only two of the six measures are based on a clear conceptual framework (Landry et al. 2001a,b, 2003; Amara et al. 2004). Third, there are limitations in the measurement approach of existing measures. Most measures use only a single item (Amara et al. 2004) or a small number of items (Landry et al. 2001a,b, 2003; de Goede et al. 2012) to assess each type of research use (e.g., conceptual, tactical). Existing measures also suffer
from mono-methods bias (Cook and Campbell 1979), exclusively relying on self-report scales, interviews, or document analysis, rather than a combination of methods (e.g., Armstrong et al. 2014). Finally, existing self-report measures and interviews ask policymakers about their use of research in general or over extended periods of time (e.g., 5 years; Amara et al. 2004), rather than in relation to recently developed policy documents or programs. This lack of reference to a specific time or policy document has been shown to impact on recall (see Wattenmaker and Shoben 1987; Walker and Hulme 1999). Zardo and Collie (2014) address this limitation by identifying research cited within a specific policy document. However, their measure does not account for uncited research that contributed conceptually, or in other ways, to development of the policy (i.e., conceptual use; Weiss and Bucuvalas 1980b). Similarly, Armstrong and colleagues’ (Armstrong et al. 2014) measure asks respondents to indicate the extent to which a range of evidence sources influenced a specific local government policy, but does not specify the type of influence.

SAGE: A new measure of research use in the development of health policy

SAGE was developed to overcome the limitations of existing measures and meet a need within the SPIRIT trial. It consists of analysis of a recent policy or program document (produced in the last 6 months) and a semi-structured interview with a policymaker who contributed significantly to the document. Based on the SPIRIT Action Framework, the interviewee is asked to describe (1) whether or not research was used to inform the development of the policy document; (2) how this research was searched for, obtained, appraised, and/or generated (i.e., research engagement actions); (3) how this research informed the development of the document (i.e., types of research use), and (4) barriers that impacted upon the use of research to inform the document. Barriers as opposed to facilitators were examined in order to provide potential explanations as to why research was not used.
1. SAGE interview developed to address limitations of prior measures and cover the key domains of the SPIRIT Action Framework

2. SAGE interview piloted and refined

3. Concrete indicators of each SAGE domain were identified using analysis of literature and interviews with policymakers

4. Indicators in each domain were categorised into key actions

5. The key actions used as the main items in the SAGE scoring checklist

Figure 2. The steps in the development of the SAGE interview and scoring checklist.

The objective of the work reported in this article was to develop the SAGE interview guide and an accompanying checklist for scoring each of the SPIRIT domains from the interview transcript and policy document. We now describe the methods and results of the qualitative analysis undertaken to develop the SAGE interview guide and checklist.

Method

Development of the SAGE interview guide

Development of the SAGE interview guide is summarized in Fig. 2 (Steps 1–2) and described below.

Generation of items for the SAGE interview guide

In order to address the domains of the SPIRIT framework, it was essential to define key terms and domains (Table 1). We used the SPIRIT Action Framework (Redman et al. 2015), seminal research on knowledge translation, and Haynes et al.’s (2014) review of health policy definitions. Based on these definitions, interview items were either adapted from existing measures or generated by the research team.

Consultation within the SPIRIT investigator team (which included policymakers and researchers with expertise in health policy, knowledge translation, research methods, and epidemiology) was undertaken to confirm the content domains to be measured by SAGE, determine whether the interview questions were likely to elicit information required to assess each domain, and check the clarity and appropriateness of item wording.

Piloting and refinement of the SAGE interview

Feedback on the initial set of SAGE interview questions was sought from a convenience sample of senior Australian policymakers and researchers in health policy to determine if the items were appropriately worded, applicable, and comprehensible. To further evaluate the face validity and clarity of the interview, pilot interviews were undertaken in three Australian health policy agencies. Participants were asked to reflect on the interview questions and process during and after their interview. Following piloting, some minor changes to SAGE were made including removal of redundant items, ensuring questions were more open-ended to allow greater exploration of responses, adding items to further understand the intended purpose and audience of the policy, and ensuring questions were phrased in reference to the specific policy document in question.

Development of the scoring checklist

In order to derive quantitative measures from the SAGE interview and documents, we needed a scoring system that could be used to efficiently and systematically code free text. The coding system needed to encompass the range of actions policymakers might perform when using research to develop a policy document. It was decided to develop a checklist that broke down each of the measured domains of the SPIRIT framework into its key subactions. If a key subaction was identified in the SAGE interview transcript or policy document, it would be ticked off the checklist. These subactions were added as prompts in the SAGE interview, to ensure interviewees provided sufficient detail to capture their actions. The steps for producing the checklist are described below and summarized in Fig. 2 (Steps 3–5).

Identifying indicators of each measured domain

With the definitions of each domain in mind (Table 1), we conducted a detailed analysis of the (1) extant literature on knowledge translation and (2) SAGE interviews with health policymakers, to obtain a diverse and broad range of indicators that represented each measured domain. Indicators were defined as concrete examples of specific actions that related to each of the key domains. For example, an indicator of appraising relevance would be evaluating whether the research findings could be applied to the target population.

Our analysis began with the knowledge translation literature in order to identify the breadth of indicators that exist in real-life policymaking, as well as definitions and operationalizations of each domain. We searched SCOPUS using specific search terms such as Research or Evidence, combined with Health policy. The abstract of each article was examined to determine if it discussed the use of research evidence in health policy. Reference lists of relevant articles were checked to identify additional papers. Indicators of each type of research engagement action, research use, or barrier were extracted from papers and tabulated according to the relevant domain, with their source references listed. Indicators for each domain were tabulated separately (i.e., 13 tables in total).

We then examined 65 SAGE interviews with policymakers from six health policy agencies participating in the SPIRIT study. The purpose of this step was to validate whether the indicators identified from the literature were relevant, appropriate, correctly worded (e.g., to match the language used in the interviews), and could be observed in the interview data. Our analysis of interviews helped refine the existing set of indicators and identified additional indicators that did not emerge from the initial literature review.

Categorizing the indicators of each domain into key subactions

The goal of this stage of the analysis was to identify the key subactions that capture the essential components of each SAGE domain.
<table>
<thead>
<tr>
<th>Measured domains in SAGE</th>
<th>Definition</th>
<th>Identification in literature review</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(A) Research engagement actions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access research</td>
<td>Searching for or otherwise identifying research to inform policy/programs</td>
<td>(Kothari et al. 2009)</td>
</tr>
<tr>
<td>Searching for research</td>
<td>The method used by the policymaker to search for research to inform the development of the policy</td>
<td>(Dobbins et al. 2001, Dobbins et al. 2007; Evans et al. 2013; Whitehead et al. 2004)</td>
</tr>
<tr>
<td>Types of research found and used</td>
<td>The types of research that were found and used by the policymaker to inform the development of the policy</td>
<td>(Dobbins et al. 2001; Evans et al. 2013; Ritter 2009; Whitehead et al. 2004)</td>
</tr>
<tr>
<td>Appraising research</td>
<td>Evaluating the quality of research and the generalisability and reliability of research results, including the applicability of identified research to local policy/program needs.</td>
<td>(Kothari et al. 2009)</td>
</tr>
<tr>
<td>Appraising relevance</td>
<td>Assessing whether recommendations, options, or interventions described in a piece of research, is applicable, compatible, or pertinent to the current policy issue</td>
<td>(Oxman et al. 2009; Wang et al. 2005; C. Weiss and Bucuvalas 1980a; C. H. Weiss and Bucuvalas 1980b)</td>
</tr>
<tr>
<td>Appraising quality</td>
<td>Assessing the scientific quality, validity, or standard of the research based on a number of factors such as its methodology, rigour, validity or credibility.</td>
<td>(Cook and Campbell 1979; Lewin et al. 2009; Mays et al. 2005; C. H. Weiss 1979; C. Weiss and Bucuvalas 1980a)</td>
</tr>
<tr>
<td>Generating new research</td>
<td>Commissioning, collaborating in or undertaking new research or new analyses to inform policy/programs</td>
<td>(Evans et al. 2013; Lemay and Sa 2014; Ross et al. 2003)</td>
</tr>
<tr>
<td>Interacting with researchers</td>
<td>Interaction, collaboration and communication with researchers through events, projects, networks, committees, etc.</td>
<td>(Campbell et al. 2009; Dobkins et al. 2002; Invaer et al. 2002; Mirzo et al. 2012)</td>
</tr>
<tr>
<td><strong>(B) Research use</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instrumental use of research</td>
<td>Use of research to directly develop content or direction of policy/programs</td>
<td>(Amara et al. 2004; Beyer 1997; Lavis et al. 2002; Milat et al. 2013; Sumner et al. 2011; C. H. Weiss 1979)</td>
</tr>
<tr>
<td>Conceptual use of research</td>
<td>Use of research to provide new ideas, understanding or concepts that influence thinking about policy/programs</td>
<td>(Amara et al. 2004; Caplan et al. 1975; Elliott and Popay 2000; Lavis et al. 2005; C. H. Weiss 1979; C. Weiss and Bucuvalas 1980a)</td>
</tr>
<tr>
<td>Tactical/Symbolic use of research</td>
<td>Use of research to justify or lend weight to pre-existing preferences and actions (Hennink and Stephenson 2006)</td>
<td>(Hennink and Stephenson 2006; Lemay and Sa 2014; Liverani et al. 2013; C. H. Weiss 1979; C. Weiss and Bucuvalas 1980a)</td>
</tr>
<tr>
<td>Imposed use of research</td>
<td>Use of research to meet organizational, legislative, or funding requirements that research be used.</td>
<td>(Amara et al. 2004; C. Weiss and Bucuvalas 1980a)</td>
</tr>
<tr>
<td><strong>(C) Barriers to research use</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal factors</td>
<td>Factors relating to the individual policymaker that may hinder his or her capacity or predisposition to use research, such as negative perceptions towards research and its use in policy, perceived or actual deficits in skills to apply research to policy</td>
<td>(de Goede et al. 2010; Helmsey-Brown 2004; Invaer et al. 2002; Oliver et al. 2014)</td>
</tr>
<tr>
<td>External factors</td>
<td>Factors outside the control of the policymaker that might influence their capacity to use research (political influences, media, deadlines, legislative/policy infrastructure, the availability of relevant research etc.)</td>
<td>(Dobbins et al. 2002; Invaer et al. 2002; Liverani et al. 2013; Mitton et al. 2007; Oliver et al. 2014)</td>
</tr>
<tr>
<td>Organizational factors</td>
<td>Organizational factors that negatively impact policymakers’ capacity or predisposition to use research in policymaking (e.g., systems to support access to research such as subscriptions to databases or journals; professional development programs)</td>
<td>(Ellen et al. 2011; Helmsey-Brown 2004; Kothari et al. 2009; Oliver et al. 2014)</td>
</tr>
</tbody>
</table>
Conceptually related indicators were grouped together in order to identify categories (representing subactions) within each domain. To do this, the first author (S.M.) grouped indicators that represented conceptually similar actions; that is, these indicators broadly fulfilled the same function or purpose, or represented a similar activity at a comparable level of complexity or sophistication. Overlapping and redundant indicators were removed or collapsed into a single indicator before being categorized into one of the key subactions. The groups of indicators were given a category label to signify the broad action they represented—these category labels became the key subactions of each domain. For example, for domain 1c: appraising relevance of research, two identified indicators were (1) evaluating whether research could be implemented cost-effectively and (2) evaluating whether research could be adapted to existing health systems. Both conceptually represented the key subaction of appraising whether research recommendations are feasible. Decisions regarding the categorization of indicators and identification of key subactions were verified by a second investigator (S.B.).

Developing the items in the SAGE scoring checklist
The key subactions that were identified following categorization of indicators formed the items in the scoring checklist. These subactions were worded in concrete terms so that they could easily be identified from policymakers’ interview responses and ticked off by an objective coder.

By grouping and collapsing indicators, we were able to identify higher-order categories (subactions) that captured and synthesized the multitude of indicators identified from the literature and interviews. This had the advantage of ensuring that the scoring checklist was both comprehensive (in that it considered all possible indicators) and more efficient to use (in that rating could focus on the key subactions rather than an extensive list of indicators).

Results
The SAGE interview
The final SAGE interview guide (Supplementary File 2) contains 22 questions that address each of the SPIRIT domains. Questions were framed in relation to the development of a specific policy or program document. Open-ended wording was used to allow the flexibility needed in an interview to account for unique contextual factors and complexities that arise when applying research in different policy contexts. Information was included to brief participants on the definition of key terms (e.g., research, appraising relevance versus quality) as contexts. Information was included to brief participants on the definition of key terms (e.g., research, appraising relevance versus quality) as conceived in SAGE. Prompts were included to help participants recall specific details such as the kinds of research they accessed, how they searched for research, or specific strategies they undertook to appraise research. These prompts were based on the subactions and examples that were identified in the qualitative analysis of the literature described below.

Identification and definition of key subactions from the observed indicators
Our analysis of the literature and SAGE interviews identified a vast number of indicators covering the SAGE domains (see Supplementary File 3 for a complete list categorized by domain and subaction). Categorization of conceptually similar indicators led to the identification of the key subactions displayed in Fig. 3. Below, we note the sources from which indicators were identified (interview, literature, or both) and provide an overview of the subactions arising from our analysis and key concepts relating to each.

Research engagement actions
Domain A1. Searching for research
A broad range of indicators was identified, primarily from qualitative studies investigating policymakers’ sources of evidence to inform policy (Supplementary File 3, Table 1). These indicators were grouped into nine subactions that took into account not only where policymakers searched for research (e.g., academic literature databases, Google), but also the deliberateness of the search (e.g., searching Medline versus finding research on one’s desk), who undertook the search (e.g., a librarian or colleague) and the rigor of the search (e.g., explicit use of key words; see Fig. 3). The subactions also incorporated novel, technology-based information sources such as social media and document sharing sites, many of which were identified from the policymaker interviews (e.g., Twitter, Research Gate).

Domain A2. Types of research found and used
Both the literature and SAGE interviews revealed the breadth of research evidence that policymakers use when developing policy (see Supplementary File 3, Table 2). Key subactions identified from our analysis related to the types of research obtained (Orton et al. 2011) and the recency of that research (Fig. 3). A third subaction—the proportion of available research that was accessed—was identified but excluded because it was considered too difficult to observe objectively. Our broad conceptualization of the types of research used by policymakers was based on Haynes et al.’s (2014) definitions of research findings. Specifically, it encompasses academic and gray literature, systematic reviews, evidence briefs, unpublished research and data, internal and external policies, books, and monographs. Defining recency of the available research was conceptually difficult because this is ultimately dependent on the policy area and whether the area is underpinned by a mature or evolving research base. We found no specific guidance for addressing this in the literature or interviews.

Domain A3. Appraising relevance of research
Very few studies explicitly addressed how policymakers appraised the relevance of research. Indicators were primarily obtained from research investigating criteria policymakers used when judging the usefulness of research. This typically incorporated attributes relating to both relevance and quality (Domain 1D). Indicators were also derived from published tools developed to help decision makers evaluate the applicability and transferability of research to a particular policy setting (e.g., Dobrow et al. 2004; Wang et al. 2005; Oxman et al. 2009; see Additional File 2, Table 3) .

Key subactions identified from our analysis related to how policymakers evaluated whether a piece of research was relevant, and the extent to which this evaluation was pre-planned or ad hoc (Fig. 3; Supplementary File 3, Table 3). Relevance was primarily judged by three factors: the degree to which research addressed the policy issue and context (e.g., target setting, population); the extent to which the research was both actionable and feasible; and the compatibility between research and the knowledge, values and experience of the policymaker or agency. Analysis of our interviews revealed that the assessment of relevance was often informal and unplanned. However, there were examples where the assessment was an explicit step in the policy process wherein policymakers either used an appraisal tool (see
Figure 3. The subactions for each research engagement action, research use type, and barrier to research use, as identified following categorization of indicators observed in the literature and SAGE interviews with policymakers.
studies investigating how decision makers evaluate the usefulness of research for decision making. Weiss and Buculavas’ (1980a) seminal study of mental health decision makers was particularly crucial here. Most indicators were drawn from published tools intended to guide assessment of the quality of systematic reviews, primary research (quantitative, qualitative, and mixed methods), or clinical guidelines (e.g., Jadad et al. 1996; Long et al. 2002; Furlan et al. 2009; Additional File 2, Table 4). Very few of these tools were designed for policymakers, so most are unlikely to reflect how policymakers typically appraise quality. Our analysis of interviews was, therefore, important for identifying real-life indicators of how policymakers appraise research quality.

Subactions identified from our analysis related to how quality was appraised, and the extent to which this appraisal was pre-planned and deliberate (Fig. 3; Supplementary File 3, Table 4). The published research primarily focused on evaluating the scientific validity of research, particularly whether a study’s conclusions were supported by the evidence and study design. In contrast, the interviews revealed that policymakers often used other factors to judge quality, such as the credibility of the researcher or institution, the clarity of the research (e.g., the aims, results, and conclusions), and its similarity to other high-quality research. Policymakers often reported seeking the advice of experts to evaluate the quality and relevance of obtained research. Furthermore, the interviews indicated that policymakers varied in the degree to which their assessment of quality was systematic (e.g., through the use of a checklist) or ad hoc.

Domain A5. Generating new research or analyses
Both the literature and interviews identified a range of research policymakers undertake to inform the development of policies. There were variations in the degree to which policymakers were actively involved in this research and the extent to which data were systematically documented and analyzed. This variation was conceptualized as subactions relating to the intensity of research generation, specifically, activities that were thorough (where findings were documented and/or analyzed and policymakers were actively involved in commissioning and/or undertaking the research, e.g., Mirzoev et al. 2012) or less intensive (where the focus was primarily on information gathering through workshops, working groups, or the like; e.g., Evans et al. 2013) (Fig. 3; Supplementary File 3, Table 5). Two additional subactions, identified primarily from our interviews, captured whether or not policymakers advocated for new research to inform the policy, and their intention to generate this research.

Domain A6. Interacting with researchers
Analysis of the literature identified many ways in which policymakers interacted with researchers (Supplementary File 3, Table 6). We grouped indicators into four subactions, the first three of which captured intensity of interaction (Fig. 3). Intensive collaborative activities typically involve formalized partnerships between policymakers and researchers, and ongoing and direct contact (e.g., mutually formulating and conducting a research project to inform policy). Less intensive interactions involve more informal or irregular contact with researchers (e.g., inviting a researcher to present their findings in a one-off meeting, having ongoing contact with researchers by phone or email). The final level captures sporadic contact (e.g., a single phone call). This categorization aligns with the three levels of collaboration proposed by Beyer and Trice (1982) (i.e., awareness, communication, and interaction) and policymaker involvement in research proposed by Ross et al. (2003) (i.e., formal supporter, responsive audience, and integral partner). The fourth subaction, identified primarily from the interviews, was the extent to which the interaction with researchers was initiated by the policymakers.

Research use
Domain B1. Instrumental research use
The literature and interviews revealed a diversity of policies that could be influenced by research. Across most policies, however, it was found that research could influence two main types of decisions: (1) the decision to prioritize (or deprioritize) particular issues, and/or (2) the decision about what action should be taken to deal with the identified issue (Fig. 3; Supplementary File 3, Table 7). Furthermore, the analysis revealed that there could be variations in the extent to which research influences these decisions. Specifically, research (1) may directly influence the core or primary aspects of a particular course of action or strategy; (2) provide additional details to a predefined course of action; or (3) have a negligible impact on policy decisions. An additional subaction that emerged from the interviews was whether research was directly cited within policy documents, particularly where specific decisions or recommendations were proposed.

Domain B2. Conceptual research use
The literature provided a rich source of indicators of how research could impact upon policymakers’ understanding of policy issues. These indicators were categorized into subactions reflecting the aspect of the health issue informed by research and the extent to which research influenced understanding (Fig. 3; Supplementary File 3, Table 8). Specific aspects of health issues that could be informed by research include general background (e.g., causes, risk factors, and current rates), the current policy context (e.g., the target setting, population, and priorities for action or neglected issues), alternative perspectives or strategies to target the health issue, or the value of research to policymaking. Similar to instrumental use, the indicators showed that there were variations in the extent to which research could improve one’s understanding of the health issue, either by informing one’s core understanding of the issues or clarifying and reinforcing the policymakers’ existing understanding of the issue.

Domain B3. Tactical research use
Many indicators were identified relating to how policymakers use research tactically (Supplementary File 3, Table 9). Subactions derived from our analysis reflected the level at which research was used to influence the views and behavior of relevant stakeholders (specified at three levels), and the types of stakeholders that policymakers sought to influence (see Fig. 3). At the lowest level, research is used primarily to inform stakeholders about key issues relating to the particular policy. At the next level, research is used as a means of justifying or backing up an established position, either to oneself or to key stakeholders. At the highest level of tactical use, research is used to persuade stakeholders to support or act upon an existing decision, or change their attitudes and behavior relating to a
predetermined course of action (e.g., convincing service providers to adopt new clinical guidelines). Examination of the literature also revealed that policymakers used research to influence a wide range of stakeholders. Stakeholders fell into two broad categories: stakeholders directly targeted or affected by a policy decision (e.g., service providers, ministers); and peripheral stakeholders that were not directly affected, but nonetheless have an interest in the policy issue (e.g., interest groups, media, civil society organizations).

Domain B4. Imposed research use
Imposed research was explored to a limited extent in the literature, whereas our interviews were an insightful source of indicators (Supplementary File 3, Table 10). Subactions arising from analysis of these indicators focused on the varying degrees to which research use can be imposed by organizations (Fig. 3). This was separated into three levels. Firstly, organizations may encourage research use, but not require it. At the next level, research use is expected by the organization, implicitly viewed as best practice, or regarded as the way the organization does its business. Importantly, however, research use is not mandated. At the highest level, research use is required and mandated, often by the organization providing a ‘policy on policy’ with firm guidelines to incorporate research during policy or program development.

Barriers to research use
Analysis of the extensive research on the barriers and facilitators to research use identify many barriers reported to impact upon policymakers’ ability to use research in policy. We broadly categorized as internal, external, and organizational barriers (Fig. 3; Supplementary File 3, Tables 11, 12 and 13).

Domain C1. Individual-level barriers
These were factors residing within the policymaker that impacted upon his or her ability to use research, such as skills at accessing, appraising, and applying research to policy, and the subjective value placed on research and its application to policy.

Domain C2. External barriers
These referred to external factors that impacted upon policymakers’ use of research such as opinions of stakeholders and political figures, and the availability of relevant (i.e., applicable, actionable, and/or feasible), reliable, and suitably presented research on a particular policy issue. A barrier identified in almost all of the interviews was insufficient time to adequately search for, appraise, and apply research to policy. An absence of time was often attributable to other external, internal, and organizational barriers (e.g., lack of access to research databases, perceived deficits in research skills such as efficient ways to search for research).

Domain C3. Organizational barriers
These barriers represented organizational factors that often greatly impacted upon policymakers’ use of research including the availability of tools and systems to support research use, or an organization culture that values evidence informed decision making.

9. Discussion
This article describes the qualitative analyses undertaken to develop a new measure, SAGE, designed to explore the ways that research is used in the formulation of a specific policy product. The specific aims of the analyses were to develop the SAGE interview guide (Supplementary File 2) and the accompanying checklist for scoring each of the SPIRIT domains (see, e.g., in Fig. 4, and the full checklist in Supplementary File 4). These practical tools are undergoing

---

A1: Searching for Research

Did the individual:

- Search academic literature databases or systematic review databases?
- Consult research experts, librarians, or reference groups to help search or identify research?
- Search grey literature sources?
- Look through reference lists of research document, citation databases, or reference manager databases?
- Use research that was on-hand, in one’s awareness or provided by colleagues?
- Use generic search engines or social media document sharing sites?

Figure 4. Example of SAGE scoring checklist for A1: searching for research.
further development and testing to gather evidence about the validity and reliability of SAGE as a measure of research utilization in policy. The detailed analyses underpinning this development work synthesizes a substantial body of research on evidence-informed policy with data from interviews with policymakers. This synthesis makes a broader contribution to our understanding of the vast range of research engagement actions, types of research use, and the barriers that negatively impact upon policymakers’ capacity to use research. Ultimately, the identification of concrete indicators and subactions enabled us to operationalize the SPIRIT framework. Our findings may have similar utility for those seeking to develop other measures. In the discussion that follows, we consider the advantages and potential applications of SAGE, the next steps in its development and testing, and potential limitations.

What does SAGE add?

SAGE has a number of features designed to address limitations of previous measures. First, SAGE is based on an explicit conceptual framework (i.e., the SPIRIT Action Framework (Redman et al. 2015)) and so it is designed to comprehensively capture not only the extent of research use but also the factors underlying research use, and the influence of the wider context in which policymaking occurs. Second, SAGE interviews focus on a recently developed policy document. Use of a concrete, context-specific reference, such as this, typically enhances recall (Wattenmaker and Shoben 1987) and should enhance the reliability and validity of the assessment of research use. It also makes SAGE a suitable measure for evaluating changes in policymakers’ engagement with and use of research over time, for example, when evaluating the impact of organization-wide interventions and programs to improve research use capacity among staff (e.g., SPIRIT, The CIPHER Investigators 2014).

The use of both interview and document analysis in SAGE provides an integrated, holistic approach to assessing research use in health policy (Hanney et al. 2003) that is typically more valid than the use of a single method. Hanney et al. (2003) recommended the use of semi-structured interviews because of their flexibility to account for the local policy context, subtle nuances and circumstances surrounding the development of the policy, and being structured enough to cover the wide range of key issues, research actions, domains, facilitators, and barriers relating to research use in sufficient depth (Boaz et al. 2009). Therefore, interviews support richer, more in-depth, and meaningful insights into how research is used when developing policy (Boaz et al. 2009).

Nonetheless, interviews alone may lead to inaccurate and/or incomplete reporting due to problems with recall and social desirability effects. In order to deal with these potential problems, SAGE incorporates numerous prompts throughout the interview based on the identified subactions to enhance policymakers’ retrieval of details relating to how they engaged with and used research. Furthermore, SAGE supplements interviews with documentary analysis to verify the veracity of interview responses (i.e., triangulation; Boaz et al. 2009), as well as to provide a context for the measurement of research use, which can then be explored in greater detail within the interview (Boaz et al. 2009). Documentary analysis could also be used to examine consistencies between the policy and the body of research relating to the policy issue, and to identify the types of research that policymakers drew upon to inform the document. However, only directly cited research would be accounted for by this method (Hanney et al. 2003). The interview, therefore, augments documentary analysis by shedding light on research that indirectly, or subtly, contributed to the document (i.e., conceptual use).

Policymakers often utilize a broad range of evidence sources besides research, such as stakeholder feedback and preferences, their own experience, opinions of experts and eminent colleagues, local and state-wide best practices, and local information (e.g., contextual details and resource availability; Black 2001; Helmsley-Brown 2004; Orton et al. 2011). SAGE goes some way toward evaluating these other types of evidence by inviting respondents to describe the context surrounding the policy, in particular, other sources of information that contributed to policy development, and the overall importance of research in this process. This indicates that SAGE can be used to provide a comprehensive and holistic assessment of what evidence (besides research) was drawn upon to inform policymaking.

SAGE aims to comprehensively measure current research use as a discrete event and the process by which research was accessed, appraised, and generated. This information may help policymakers and organizations identify possible ways to strengthen research use in policy development. For example, using the interview with the checklist may reveal that policymakers within an agency were not searching for research by using academic literature databases or consulting experts to help locate relevant research. This may assist organizations to identify the need for targeted professional development, for example, workshops on using bibliometric databases (e.g., training workshops; Ellen et al. 2013), or establishing stronger links with researchers and other experts (Kothari et al. 2005). Alternatively, it may point the agencies toward increased investment in systems and tools to support efficient access to research (e.g., providing access to databases, librarians, and other research experts; Ellen et al. 2011). SAGE can then be used as a measure of change to determine whether these mechanisms were effective in improving research engagement scores of policymakers, and whether improvements in scores are associated with gains in wider economic and health outcomes.

Finally, the scoring checklist for SAGE enables quantitative measures to be derived from rich qualitative data. The checklist is designed to support accurate and reliable coding of policymakers’ use of research on the basis of their responses to the SAGE interview and the accompanying policy document. Such quantitative ratings are particularly important for assessing whether investment in efforts to support the use of research in policy have had a measurable impact.

Potential limitations of SAGE

SAGE has been designed to be a general measure of research utilization over a range of policy issues and contexts. However, every context and policy issue is different. Certain research engagement strategies, types of research use, and barriers to use may be more appropriate in particular contexts than others. For example, the development of guidelines may best be informed by searching academic databases and retrieving systematic reviews and peer-reviewed literature, whereas implementation or workforce planning policies may best be informed by internally produced research, evaluations, and policies from nearby jurisdictions (i.e., gray literature). As a result, SAGE may not be an appropriate measure of research utilization in all policy contexts. Ongoing validation of SAGE will be required following its development to determine in which contexts, settings, and for which policies it can be used as a reliable and valid measure of research use.
A second limitation is that SAGE addresses barriers, but not enablers, to research use in policy. Barriers were the focus in order to provide explanations as to why research was not optimally used, or engaged with by staff. However, the SAGE interview does in fact ask respondents to describe what factors enabled research use. Similarly, many of the barriers in the checklist (Supplementary File 3, Tables 11–13) are also enablers to research use. Consequently, the checklist could easily be adapted to score enablers mentioned in the interview.

In addition, although SAGE can identify policymakers’ research engagement actions, research use, and barriers to research use, it does not assess the availability of specific organization-wide structures, mechanisms, and systems that may improve policymakers’ capacity to use research. In addition, although SAGE can identify whether individual policymakers experience barriers to research use, it is not designed to identify overarching organizational structures that may contribute to these barriers. For example, a policymaker may report a lack of research appraisal skills in a SAGE interview. However, this may be symptomatic of an organization that lacks a supportive culture and appropriate systems and supports to facilitate research use.

These gaps are addressed by other measures in the suite of instruments developed to operationalize SPIRIT. The first of these is ORACLe (Organisational Research Access Culture and Learning; Makkar et al. in press; The CIPHER Investigators 2014), which specifically measures the extent to which organizations possess different tools, supports, and resources that build capacity to access and apply research to policy. A related measure entitled SEER (Seeking, Engaging with, and Evaluating Research; The CIPHER Investigators 2014) assesses policymakers’ perceptions regarding the value of research to policymaking, their knowledge and skills in applying research to policymaking, and the availability of tools and processes to support research use within their organization. ORACLe and SEER should be used alongside SAGE to capture individual-level and organization-wide barriers, and to help identify specific initiatives to improve the organizations’ research use capacity.

A number of practical issues arose during the piloting phase. For example, the time required to complete the interview was an issue, with durations ranging between 30 and 90 minutes. Such a lengthy interview might not be feasible in busy policymaking organizations. As described in the method, it was important to ensure questions were ordered logically and worded appropriately to prevent repetition from both the interviewer and interviewee, and enhance the overall efficiency and flow of the interview. Other feedback included the importance of defining all the key terms and providing concrete examples to ensure valid responses are elicited. For example, if interviewees are not told that research includes gray literature, their responses may be restricted to the use of academic literature, which would lead to invalid responses. A final key issue was the importance of ensuring that interview questions were flexible, and could be adapted to account for the broad range of policy documents that can exist. Although these issues were resolved at the pilot stage, other issues are likely to emerge with further testing of SAGE in different policy contexts and organizations. Therefore, ongoing evaluation of the practical utility of SAGE is necessary.

**Further development and testing of SAGE**

Two essential next steps for SAGE are: (1) further developing the SAGE scoring tool so it can generate valid total scores for each

### A1: Searching for Research

**Did the policymaker...**

- Search **academic literature databases** or systematic review databases? ☑️
- Consult **research experts, librarians, or reference groups** to help search or identify research? ☐
- Search **grey literature sources** (e.g., local intranet, authoritative websites, agency websites, grey literature databases or trials)? ☑️
- Look through **reference lists** of research document, citation databases (e.g., Web of Science), or **reference manager databases** (e.g., EndNote)? ☑️
- Use research that was **on-hand**, found on his/her computer/bookshelf/desk, in his/her awareness, found through **email alerts, conferences**, or provided by **colleagues**? ☐
- Use **generic search engines** (e.g., Google) or **social media document sharing sites** (e.g., Research Gate)? ☐

**Score**

<table>
<thead>
<tr>
<th>9</th>
<th>2.83</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>1.56</td>
</tr>
<tr>
<td>7</td>
<td>1.42</td>
</tr>
<tr>
<td>6</td>
<td>1.22</td>
</tr>
<tr>
<td>5</td>
<td>1.08</td>
</tr>
<tr>
<td>4</td>
<td>0.88</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Total** 2.83 + 1.42 + 1.22 = 5.47

*Figure 5. Example of SAGE scoring checklist with weighted point values for 1A: searching for research.*
research engagement action and type of research use (which is now complete); and (2) evaluating the reliability and validity of SAGE.

Firstly, the subactions derived from the qualitative analyses (as shown in Fig. 2 and Supplementary File 3) have been used to create a checklist-based scoring tool that will allow external coders to rate policymakers’ level of research engagement and use in the development of a policy document (see Supplementary File 4). Using the SAGE interview and the accompanying policy document, raters mark whether or not policymakers engaged in each of the key subactions listed under the research engagement action and research use domains. A score is then assigned to each subaction and a total score calculated for each measured domain. The score attributed to each subaction was determined in a separate study in which experts with extensive knowledge and experience in health policy and translation research completed a choice survey in which they rated the relative importance of each subaction for undertaking evidence-informed policy (Makkar et al. 2015a,b). A weighted score was calculated for each subaction based on the expert rating. Total scores for each of the measured domains are calculated by summing the scores for all ticked-off subactions in the scoring checklist. See Fig. 5 for an example on how a total score for ‘searching for Research’ would be obtained from interview responses.

The information regarding barriers from Fig. 3 has been used to develop a comprehensive barriers checklist (Supplementary File 5). The checklist requires external coders (or the policymakers themselves) to rate the extent to which each barrier negatively impacted the policymaker’s ability to use research to develop the policy document.

Having developed this scoring system, we are currently testing the inter-rater reliability of SAGE, specifically by examining the level of agreement between coders on each of the measured domains. We are testing validity by assessing inter-rater agreement on SAGE between trained coders and experts in health policy and research, and aim to determine whether SAGE scores are predictive of greater research use in policy development.

Conclusion

In this paper, we have described the qualitative analyses undertaken to develop a comprehensive measure of research use in health policymaking. This primarily involved explicating the key elements and features of each type of research engagement action, research use, and barrier to research use. This information was obtained through detailed analysis of the current literature on evidence-informed health policymaking and qualitative analysis of over 60 interviews with policymakers from a diverse range of public health organizations in Australia. With this information, we have developed a preliminary scoring tool to assess the research engagement actions undertaken by a policymaker, the ways research was used by the policymaker, and the barriers that impacted upon the their ability to use research. By measuring research use validly, policy organizations may be able to implement initiatives to improve research use among staff, leading to evidence-based policies, more efficient health spending, and better health outcomes for the community.

Supplementary data

Supplementary data is available at Reeval Journal online.

References


