10-1-2012

Improving Language Policy and Planning Through Evaluation: Approaches to Evaluating Minority Language Policies

Haley De Korne
University of Pennsylvania

Improving Language Policy and Planning Through Evaluation: Approaches to Evaluating Minority Language Policies

Improving Language Policy and Planning Through Evaluation: Approaches to Evaluating Minority Language Policies

Haley De Korne

University of Pennsylvania

An increasing number of regional, national, and international policies promote the use of minoritized, Indigenous languages in education and other social domains, but little is known about how successful these policy approaches are. There is a lack of research on minoritized language (ML) policy development and implementation, with no established procedure for evaluating the success of diverse policies. This paper aims to contribute to the on-going development of ML policy evaluation (and language policy evaluation in general), with an emphasis on the eventual use of the evaluation towards policy improvement. Case studies of three different ML policies are analyzed to determine: What policy evaluation approaches have been employed? And, how effective were these evaluations in improving the policies? The resulting discussion considers promising ways of assessing the outcomes of policies in complex ML contexts, including multiple research methods, on-going evaluation, multi-genre and multilingual dissemination of results, and participation from stakeholders at all levels of policy implementation.

Introduction

There is a movement, driven both locally and globally, by institutions and individuals, to improve the social status and vitality of some of the languages that have been minoritized through processes of political and economic colonization (cf. Hinton & Hale, 2001; UNESCO, 2010). An increasing number of regional, national, and international policies promote the use of minoritized Indigenous languages in education and other domains (De Korne, 2010; Hornberger, 2005). A variety of policies have attempted to promote the inclusion of minoritized languages (MLs) in education in order to improve minority student achievement and educational equality, but little is known about how successful these policy approaches are. Language policy commonly refers to official documents created by governments or other authorities; however, it is also constituted by the practices of stakeholders involved in developing and implementing these texts (cf. Canagarajah, 2005; Menken & García, 2010). This working paper is limited to an examination of ML policies at the level of legal texts, but attempts to discuss social practices around these texts through consideration of evaluation and policy improvement initiatives.

Language policy in ML contexts may have multiple goals, including the provision of better overall education outcomes for children from minority language backgrounds, (some degree of) literacy or oral proficiency in minority

languages, and/or awareness of the history and culture of a minority language group. Stakeholders may have different goals and agendas in relation to the same policy. Factors that impact the outcome of ML policies are also numerous and complex, including education funding, shifts in population, dialectal variation and trends of language use (Bratt Paulston & Heidemann, 2006; García, 2009). The multiple goals, expectations, and intervening variables make it difficult to arrive at substantiated evaluations of a policy's outcome(s). Additionally there is a lack of research on ML policy development and implementation (Grin, 2003), with no established procedure for evaluating the success of diverse policies.

This paper aims to contribute to the on-going development of ML policy evaluation (and language policy evaluation in general), with an emphasis on the eventual use of evaluation towards policy improvement. Case studies of three different ML policies, listed in Table 1, are analyzed to determine: What ML policy evaluation approaches have been employed by political authorities? How were these evaluations used to improve the policies? Each policy case study considered here was evaluated differently by the authorities upholding it, allowing for comparison of the strengths and weaknesses of different evaluation models and their contributions to the potential effectiveness of the policies. The resulting discussion considers how political authorities can assess the outcomes of policies which they implement in complex ML contexts, emphasizing evaluations that lead to improvements.

Research on language policy takes many forms, including classic typologies of policy types and steps (cf. Cooper, 1989; Haugen, 1983), analyses of language orientation (Ruiz, 1984), ethnographic implementation studies (cf. Canagarajah, 2006; Johnson, 2009; Skilton-Sylvester, 2003), historical analyses (Spolsky, 2004), and recently cost-benefit and impact analyses (Grin, 2003). As with all social research, there are different approaches advocated by scholars with different epistemological orientations. For example, coming from the ethnographic tradition Hornberger and Hult (2008) propose an ecological approach to understanding language policy by collecting "multidimensional data" (p. 285) on how the policy relates to individuals, to local contexts, and to the promotion of equality among language groups. From a sociological perspective, Tollefson (1991) proposes a "historical-structural approach" in order to "relate language policy to broader issues of economic development and sociopolitical change" (p. 32).

While these approaches to understanding language policy offer valuable insights, they differ somewhat from work in the field of applied policy and program evaluation that aims explicitly to assess and recommend improvements to a social policy. The use-oriented nature of evaluation is one of its fundamental features and is largely responsible for setting the field of evaluation apart from the description-oriented social sciences (Leeuw & Vaessen, 2010). Increasing demand for accountability and evidence-based policy currently drives research in many areas of social policy, including education. Unlike some forms of descriptive research, evaluative research is typically intended to provide results that can be used to impact the topic that has been explored, by making recommendations on the basis of positive or negative judgments. Although evaluation was introduced as a step in language planning by Rubin (1971) and was subsequently included in Haugen's (1983) and Fishman's (1980) typologies of "language planning steps," this step has received relatively less scholarly attention than issues of agendas, codification and

implementation. In a contested area such as the quality and equality of education for minority language communities, there may be multiple desired outcomes that an evaluation could attempt to measure, including economic factors, stakeholder perceptions of quality, and empirical measures of language proficiency. Resulting recommendations could likewise focus on improving any of these areas, including policy inputs, supporting variables, and/or specific desired outputs.

The work of François Grin (2003; Grin & Vaillancourt, 1999) is notable in that it combines some traditional academic concerns with an applied evaluation focus on economic and other quantitative outcome measures. Rather than focus only on language proficiency measures, Grin (2003) promotes measures of "effectiveness," "cost-effectiveness," and "democracy" (or "the democratic character of the procedures of the selection, design and evaluation of policies") as tools for ML policy evaluation (p. 89). He also explicitly states that economic factors should be only one criteria among many in the evaluation of a policy, noting that all policy choices involve political debate and negotiation, and thus cannot be decided by quantitative factors alone (Grin & Vaillancourt, 1999). He notes meanwhile that economic factors are especially persuasive to policy-makers, and thus valuable to consider. Grin's work, bringing economic theory to bear upon ML policy evaluation in order to make the evaluations more useful to policy-makers, is highly regarded by other scholars in the field (e.g., Byram, 2008). Regardless of which factors decision-makers weigh most heavily, it is clear that considering a variety of factors is beneficial to both understanding and evaluating a social policy.

Using Language Policy Evaluations

Evaluations can be labor-intensive and costly, and in a domain such as ML education where funds are scarce it is important that such an activity have significant value, and not become an exercise in useless data production. There are numerous perspectives from which to analyze and critique the strengths and weaknesses of different evaluation approaches, or in other words to evaluate an evaluation model. These include appropriateness, cost-effectiveness, quality and rigor of data, and utility or impact, all of which cannot be considered within the limited scope of this paper. The factor that will be emphasized here is the model's potential for impact.

An evaluation that results in tangible changes to a policy text or practice can be considered to have a direct use towards policy improvement. An evaluation that does not have a "direct use" (e.g., through concrete recommendations adopted) may still have a "conceptual use" (e.g., through influencing awareness on the topic) (Rossi, Freeman & Lipsey, 1999, p. 432). Weiss (1980) has noted that the conceptual use of evaluation, or *knowledge creep* can be significant because positions and information on social phenomena (and even definitions and categorization of these phenomena) may influence a wider audience's perception of these issues. Pawson (2006) argues in support of this, stating that "the influence of research on policy occurs through the medium of ideas rather than of data" (p. 169). Nevertheless, there are also numerous examples of direct use of evaluation results, although they may not always be immediate (cf. Leviton & Boruch, 1983). Whether or not an evaluation has resulted, or will result in use is thus a complex issue, and cannot be determined by obvious evidence of direct use alone. For example, descriptive

language policy research may also result in knowledge creep, even if such research does not explicitly advocate changes in social practice.

In addition to determining whether an impact has occurred as the result of an evaluation, the subsequent goal of this line of research is to design evaluations with a future impact. With this in mind, Leviton and Hughes (1981) synthesize factors affecting utilization of an evaluation as: relevance; communication between researchers and users; information processing by users; plausibility of research results; and user involvement or advocacy. These factors indicate that planning for evaluation impact requires careful consideration of the policy context. When policies aim to increase educational achievement and equality for minority language students, participants' perspectives (vis à vis equality and achievement) are essential to the thorough evaluation of these policies, and to acceptance of evaluations. Hatry and Newcomer (2004) propose credibility as one of the "touchstones of methodological integrity" for evaluation, noting that "evaluation findings are more likely to be accepted if the program stakeholders perceive the evaluation process and data to be legitimate" (pp. 548-549). In relation to the implementation of findings, Rose (1993) also notes that "failure to take into account the values of the dominant coalition in government will leave a lesson in limbo; it can be applicable but, if politically unacceptable, it will not be applied" (p. 15). ML policy analysis that seeks to alter and improve policy implementation must therefore maintain close consideration of stakeholders' perspectives and the social ecology of the policy context. Results must also be distributed to stakeholders if knowledge creep impacts are to occur. In the context of ML policy, not only the medium of distribution (e.g. scholarly article, government report, publicity pamphlet) but also the language that results are distributed in will influence the reach of the recommendations. For example, the reports of the policy evaluations profiled below were all published in majority languages, and not in the target minority languages concerned. Publishing them in minority languages would change the audience of the reports and allow them to circulate among different stakeholders.

In moving forward towards use-oriented evaluations of minority language policy, it seems likely that contributions from a variety of epistemological and methodological positions will be welcome. In other words, the ethnographic considerations for stakeholders' insights and local contextual detail, and the sociological considerations for historical-structural constraints do not need to be abandoned in favor of economic measures alone; rather, each approach brings a valuable dimension to the improvement of ML policies.

Case Studies of ML Policy Evaluation

Despite the lack of consensus on methods of evaluating ML policies, as ML policies have increased in number during recent decades several have been evaluated by the government bodies that established these policies. The three ML policy evaluations considered here are all full-coverage programs, meaning that they apply to all members of a population sample. This eliminates the potential for certain kinds of evaluation designs, such as randomized interventions where the impacts are compared between contexts of policy coverage and non-coverage. Instead these evaluations tend to rely on what Rossi et al. (1999) term "shadow controls" (p. 356), or evaluation by program stakeholders and/or experts, rather

than the quasi-experimental, control or comparison studies that are favored in medical and some areas of social policy research. This approach may be a result of the relatively small scale of these programs and their target populations.

It is important to note that although these three ML policies share a common aim of promoting MLs in education and other domains, they vary in their intervention strategies (see Table 1). The policies also vary in the size of population covered and the degree of legal strength that they have, and exist in different educational environments. This is therefore not a direct comparison of different approaches to evaluating the same policy. Instead it is an attempt to analyze each evaluation model in relation to the specific policy that it evaluates, while establishing a few points of comparison and analyzing the general merits of the evaluation methods. There are distinct differences that can be compared across the evaluation approaches employed in each case, especially in regards to the frequency of the evaluation and who is included in the evaluation process, as analyzed below.

Table 1: *Three Policies that Promote Inclusion of MLs in Education: Relationship Between Policy Approach and Evaluation*

ML Policy	Washington State First Peoples' language, culture and oral traditions certification (2007)	The Northwest Territories Official Languages Act ([1988] revised 2003)	European Charter for Regional or Minority Languages (1998)
Authority(ies)	Washington State Department of Education, Washington State Legislature & Tribal governments	Northwest Territories Legislature	Council of Europe & National governments
Policy Approach	Certify more ML teachers through alternative process	Raise the legal status of MLs & increase school initiatives	Require some ML use across all social sectors
Evaluation approach	Collaborative 3-year pilot project	Government-run yearly evaluation	Independently run evaluation conducted every 3 years

With contextual contingency in mind, these three cases of ML policy evaluation will be analyzed in turn. Tables 2-4 present logic models that combine policy and evaluation logic to give an overview of how each evaluation model measures policy outcomes and outputs. The general strengths and weaknesses of each model are discussed and recommendations for improvements in the evaluation model are given. These discussions are not exhaustive; they raise only a few issues related to each evaluation approach. Then, a comparative model (Table 5) is used to consider the potential that different evaluation approaches have to impact or improve policy through direct and/or conceptual use.

Case study #1: Washington State First Peoples' language, culture and oral traditions certification (2007)

This policy aims to increase the presence of Indigenous (Native American) languages in public elementary schools and secondary schools in Washington State as a way of improving Native students' achievement in school and enhancing non-Native students' awareness of regional history and culture. The policy was established as a pilot project, with the requirement of evaluation in order to determine if and how the policy would become permanent. The evaluation model is outlined in Table 2.

General strengths and weaknesses of this evaluation approach

There was significant inclusion of stakeholders (including the Tribal governments' representatives, participating teachers and the State Board of Education Committee), resulting in greater internal validity through collecting data on diverse stakeholders' perspectives. This approach also resulted in sharing of evaluation costs across stakeholder groups. Each participating Tribal government evaluated the outputs of the program, such as the number of teachers receiving training and certification; however, only some evaluated the effects of the program on student achievement. While straightforward quantitative data can be gathered regarding the numbers of collaborations undertaken, certifications awarded, and classes taught (outputs), the evaluation of the impact on student achievement (outcome) requires further definitions of measurements and selection of methods (e.g., observations, interviews, written and/or oral language proficiency assessments). Because the program outcomes were evaluated by stakeholders at the local level, different measures were used to determine the effect of the program on student achievement in different individual reports. Tribal stakeholders and the State committee did not all conduct evaluations in the same way, making their combined final report less consistent and potentially less credible. As indicated in the final report (Washington PESB, 2007), a consensus on the outcomes of the program was established through regular meetings and consultation among stakeholder groups, or in evaluation terms, shadow controls. There may be additional concern over the validity of results that are drawn from a pilot study, rather than a full-fledged implementation of the policy.

There was a clear path for evaluation results to improve the policy by recommending changes to the pilot structure before making the policy permanent. The final report (combining individual reports from tribes and the state committee) concludes that the policy produced positive outputs and outcomes and should become permanent, but made several specific recommendations for improvement. These recommendations came out of meetings between stakeholder groups throughout the pilot period and while the final reports were being prepared, rather than from individual reports themselves. The recommendations were adopted when the policy became permanent in 2007.

Table 2: *Policy #1 Content in Relation to Evaluation Approach*

Policy #1	Policy Strategy	Evaluation Method	Means of verification
Goal	Educational success for Native & non-Native students through Native language teaching	3-year pilot project	Overall success determined by joint report from state committee & tribal representatives
Inter-vention	Tribes permitted to certify language & culture teachers in collaboration with the State	Committee conducts on-going meetings & observations with tribal stakeholders	Number of agreements signed
		Tribal stakeholders document local program implementation	Stakeholder perspectives on collaboration process
	Schools encouraged to hire language & culture teachers	Not evaluated	Not evaluated
Output	Certification of Native language teachers	State committee & tribal stakeholders end-of-pilot report	Number of certifications awarded
	Native languages taught in schools	Tribal stakeholders end-of-pilot report	Number of classes & students taught
Outcome	Improved government-to-government relations	State committee & tribal meetings; end-of-pilot reports	Stakeholders' perspectives; observation data
	Improved school-tribe relations	Tribal end-of-pilot reports	Variable/ not consistently evaluated (stakeholders' perspectives; observation data)
	Improved retention and success for Native students	State committee & tribal meetings; tribal end-of-pilot reports	Variable/ not consistently evaluated (stakeholders' perspectives; observation data)
	Improved awareness for non-Native students	Tribal end-of-pilot reports	Variable/ not consistently evaluated (stakeholders' perspectives; observation data)

Considerations for improvement

Delegating evaluation of outputs to local stakeholders appears to be efficient and effective in this case; however, the evaluation of outcomes across different contexts where no unified measures have been established is clearly problematic

when attempting to combine these evaluations in one report. It would be beneficial to collaboratively establish common procedures for measuring program outcomes, which could still be carried out by stakeholders at the local level, thus retaining the high level of participation but increasing the data consistency and quality. Explicit discussion of appropriate evaluation criteria, leading to a balance between locally relevant criteria for measuring program success and widely accepted criteria would help to bridge the divide between a central policy and local implementation. Although the evaluation resulted in direct use, improving the pilot model in the permanent policy, there was no provision for future evaluation of policy progress, either by stakeholders or other experts. This one-off evaluation model is a shortcoming, as contexts for education shift over time and require on-going adaptation of policies. This is especially true for ML policies with corrective goals such as this one. For instance, if the policy is successful in closing the gap of achievement and equality between minority and majority students, as it aims to do, then it will need to be adjusted to become a policy that maintains minority student achievement.

Case study #2: The Northwest Territories Official Languages Act ([1988] revised 2003)

This policy has the broad goal of providing equal services to speakers of all eleven official languages in the Northwest Territories (NT) of Canada. Following a special evaluation (Northwest Territories, 2003) this policy was amended, (1) to establish several government authorities (two individual posts and two advisory boards) dedicated to advancing MLs in education and other public sectors, and (2) to require an annual report on the implementation of the policy by one of these authorities, the Minister responsible for Official Languages (Minister for OL) (cf. Northwest Territories, 2009). The evaluation model is outlined in Table 3.

General strengths and weaknesses of this evaluation approach

The breadth of this policy makes it a challenge to evaluate. The broad goals of the policy (establishing equal status and opportunities for users of all official languages) reflect the complex reality of multilingual societies. Rather than targeting a specific issue to improve, which then may not have the desired results because other related issues are not addressed, this policy aims to address the problem of unequal treatment of languages and speakers on multiple fronts. However, these broad goals can be difficult to operationalize in the implementation and evaluation of the policy. For most goals no specific targets have been set, either by the Act itself or the authorities that it established. Without fixed targets such as the number and nature of services to be made available, or the level and quantity of language instruction occurring in schools, the annual reports remain largely descriptive rather than evaluative, despite being charged with providing "an evaluation of the effectiveness and efficiency of the policies and programs of government institutions relating to Official Languages…" (NT Statutes of the Official Languages Act, Section 27(1)). It may be beneficial to move forward on multiple fronts, improving all social services, but specific targets would greatly improve the potential for evaluating the efficiency and effectiveness of improvements.

Table 3: *Policy #2 Content in Relation to Evaluation Approach*

Policy #2	Policy Strategy	Evaluation Method	Means of verification
Goal	Equal use of Aboriginal languages across public sectors, including education, is ensured by government & by language communities	Annual report by Minister for OL	Success determined by reports from several government agencies
Inter-vention	Official language status upheld for 9 Aboriginal languages	Annual report by Minister for OL	Descriptive statistics on language use; Expenditure for all OL-related programs
	Administrative structure established to promote language use in all sectors, including the Minister for OL, Languages Commissioner, & 2 Boards	Official Languages Board & Aboriginal Languages Revitalization Board report on their activities to the Minister for OL	Number & nature of Boards' activities
		Languages Commissioner reports to the Legislative Assembly	Number & nature of investigations & any recommendations
Output	Government services are available in Aboriginal languages	Report evaluates progress in the delivery of services	Number & nature of services available
	Structural support for use of Aboriginal languages in schools increases	Report describes progress in teacher training programs & curriculum development	Number of participants & graduates of teacher training programs; description of curriculum
Outcome	Speakers of all NT languages receive equal treatment	On-going reporting on language services available	Number & nature of services
		Languages Commissioner's report on investigations & complaints	Number & nature of complaints
	Aboriginal languages are maintained & revitalized	On-going reporting on use of languages	Descriptive statistics on language use

The number of government authorities charged with implementing and evaluating the Act brings attention to language issues within the government due to the production and presentation of reports across sectors. The lack of

consideration for the role of community stakeholders alongside the government is a weakness, though. The NT Official Languages Act recognizes that the responsibility for its success must be shared among language communities and the government (NT Statutes of the Official Languages Act, Preamble), however little provision is made for collaborating with communities or evaluating the role of communities. The Official Languages Board and the Aboriginal Language Revitalization Board consist of representatives from each community. While they are given the power to consult and advise on OL issues, they are not guaranteed the right of consultation, nor are they required to report on progress from the perspectives of their respective communities. Ideally, the Boards can serve as a bridge between the language communities and the government authorities with decision-making power, collaborating with the Minister for OL and the Languages Commissioner to improve the policy measures. Consultation between these authorities is encouraged but not required, and thus it is not certain how each actor will contribute to eventual improvements in the policy. The position of community stakeholders, especially in regards to improving the policy, is unclear.

The regular and frequent nature of the evaluation is a strength, establishing time-series data on the policy at yearly intervals. Producing a thorough evaluation every year is also a considerable expense, and may have the unintended effect of making the reports less politically impactful because they appear so frequently. Assessing the NT Official Languages Act from inside the government presumably allows ready access to information about all government initiatives, but it also has some distinct drawbacks. It is generally not explicit which stakeholders have contributed to the reports; the voice adopted is one of governmental oversight and authority, as invested in the Minister for OL. Nor is the process of data collection made transparent, weakening the reliability of the reports. The Minister for OL is responsible both for implementing the Act, and for evaluating the effectiveness of its implementation in annual reports. It is thus not surprising that the resulting reports focus on progress made towards an undefined benchmark of success, rather than any shortcomings or areas in need of improvement in order to reach pre-established goals. It is also unsurprising that the evaluations focus on government activities, almost excluding the role of community stakeholders, despite communities' important place in the policy rationale.

Considerations for improvement

A clearly established role for community stakeholders in the implementation and evaluation of the policy would help to shift the policy away from its heavy bureaucratic structure. A basis for this already exists in the Language Boards, and could be strengthened by specifying roles to be played by community-level organizations in implementing and evaluating the policy.

In relation to the use of the available data, the reports would benefit from organizing the data they present in comparative ways. Although the frequency of the evaluation establishes time-series data on language use and program expenditure, each report is submitted independently, and generally does not provide a contextual comparison of previous data. This means that the interested reader must compare evaluations manually in order to get a sense of trends in the data. The descriptive data could be more informative if it were presented in the

context of data gathered in previous reports. Data about program participation (such as teacher training and Board activities) would also be more useful if presented in relation to established goals or expectations for successful participation.

In terms of the equality of educational services in Aboriginal languages, the reports do not open the commonly called black box of what goes on inside schools, but remain at the level of teacher training, curriculum development and the construction of school facilities or new schools. A stronger evaluation of this piece of the policy could be carried out through greater participation of education stakeholders such as administrators, teachers and parents. The evaluation does not consider whether the policy is meeting the stakeholders' needs. If the evaluation did address this concern, it would prove useful for future policy work.

Finally, the general focus on progress rather than problematic areas means that the evaluations do not draw as much attention to possible areas of improvement as they could. Bringing in stakeholders from outside of the government could help to highlight recommendations for improvement. As a report submitted directly to the Legislative Assembly, and with a history of resulting in direct use, the impact potential of this evaluation is high. However, the current method of evaluation does not result in much attention being given to areas in need of improvement, thus weakening the potential for positive impact.

Case study #3: European Charter for Regional or Minority Languages (1998)

This is a multinational policy that encourages member states (countries) of the Council of Europe to select specific targets that are appropriate to their national context in order to improve the status of minority languages in society (for example through specific increases in the use of MLs in education, government services, the media, and cultural activities). Each signatory state must submit a triennial report on the progress they have made towards their chosen targets and be evaluated by an independent committee of experts appointed by the Council of Europe. The evaluation model is outlined in Table 4.

General strengths and weaknesses of this evaluation approach

This policy aims for on-going improvement through regular evaluations, which are intended to ensure that ratified measures are carried out, and to propose improved measures once initial ones have been met. The use of external evaluators in conjunction with reports produced by state governments should increase the accountability and rigor with which policies are evaluated. However, member-states do not necessarily respond to the critique of their policy implementation, or look for guidance in improving the strength of the policy (Council of Europe Secretary General, 2009). In some cases the committee of experts has reported the same shortcomings in a state over several evaluation cycles. Evaluation results aim to guide national authorities to better policy implementation, but if national authorities do not voluntarily adhere to the evaluation results, there is no mechanism to ensure that results will be applied (European Bureau for Lesser Used Languages conference, Leeuwarden, October 2009).

Table 4: *Policy #3 Content in Relation to Evaluation Approach*

Policy #3	Policy Strategy	Evaluation Method	Means of verification
Goal	Protect & promote MLs in member states of the Council of Europe	3-year independent evaluation cycle	Report of the committee of experts
Inter-vention	Establish legal framework for the promotion of MLs	Policy creation not evaluated	Not evaluated
	Oversee the implementation of state-specific policies to promote MLs	States submit a triennial report	Government actions towards policy implementation
		Committee of experts conducts triennial investigation of each state	Observations & interviews with stakeholders; Fact-checking of state report
		Secretary General of Council of Europe submits a biennial progress summary	Synthesis of committee of experts' reports on all states
Output	States create & ratify policies	Policy creation not evaluated	Not evaluated
	States implement policies	States' triennial reports; committee of experts' triennial reports	Observations & interviews with stakeholders; Fact-checking of state report
Outcome	MLs increase in status	On-going monitoring of whether or not policy measures have been implemented as ratified	Committee of experts' perceptions
	Cultural pluralism & coexistence is achieved between official languages & MLs	On-going monitoring of whether or not policy measures have been implemented as ratified	Committee of experts' perceptions

The use of specific targets and the facilitation of evaluation by external committees lead to increased reliability, external validity and uniformity across evaluations, as committees conduct evaluations of multiple national contexts. This two-tiered evaluation system may be costly, however, and thus may not be sustainable if the funding situation for the policy changes. Additionally, the perspectives of experts (scholars from other member states, with experience in language policy) based on brief visits to the member state make up the majority of the evaluation. As the committee is hosted by the national government during these investigative visits, they may be exposed largely to what the government would

like them to see. The policy states that legally established non-profit organizations representing ML communities may submit information to the committee, which will be considered at the committee's discretion. This is a very small window through which community representatives have the potential to participate in the evaluation of the policy, and there is no guarantee that they are aware of their right to do so. The top-down nature of the evaluation and the lack of consideration for collecting data from a variety of stakeholders lead to concerns for internal validity. Without considering the impact of the policy on the status of ML speakers at the local level, many areas of potential improvement are likely to be overlooked as well. As noted in the most recent overall report on the policy, "the potential role that civil society could play in the [policy] process is far from being fully exploited. …Often [stakeholders] are not fully aware of the rights and duties created by the Charter and the way they could successfully shape both the recommendations made during monitoring and their subsequent implementation" (Council of Europe Secretary General, 2009).

Considerations for improvement

While the use of external evaluators is beneficial, the inclusion of input by stakeholders at various levels of policy implementation would help to increase the sensitivity of the evaluation instrument, and its potential to provide recommendations that will be acceptable to stakeholders. The onus is on the evaluators to elicit the assistance and input of local stakeholders, and on the policy administrators to engage them in implementing and improving the policy.

The implementation targets are chosen by the state from among numerous targets that the policy proposes (such as number and nature of services provided in MLs, level and quantity of ML instruction in schools, etc.), meaning that states are evaluated on whether or not they have achieved targets that they have individually set for themselves. Therefore they should be invested in achieving these targets, and thus the trend of states not complying with recommendations and leaving targets unmet is understandably perplexing for the evaluators. Further consideration is evidently needed to understand how the results of evaluations are received by member-states in order to increase the likelihood that national authorities will respond to the external recommendations. Considerations could include the overall evaluation approach, the nature of the data collected, how the results are presented and disseminated, and any incentives or penalties that may be appropriate. For example, if state governments do not take account of external evaluations of the pre-established targets, other metrics could be considered, such as economic factors (e.g., cost-benefit analyses) or social impacts (e.g. matching/comparison evaluations of regions or schools where policies are implemented versus those where they are still lacking).

Comparing Evaluation Impact Potential

It is not possible to compare evaluation models in absolute terms, as each is uniquely adapted to the policy that it measures. However, a comparative perspective on the potential utility of each evaluation method may help to uncover common strengths and weaknesses in these evaluation approaches. In Table 5 the

evaluations are compared in terms of probable direct use and conceptual use. While documenting direct use is relatively straightforward, it is much harder to determine whether conceptual use has taken place. Each of these evaluations has resulted in some direct use, although often not to the degree that evaluators may have wished. Whether or not they have resulted in conceptual use would require further inquiry in the local social ecology, although some basic observations can be made.

Table 5: *Impact of Evaluation Methods*

	Evaluation Method	Direct Use	Conceptual Use
Policy #1 Washing-ton State	3-year pilot study: Documents outcome of pilot project; determines whether policy will become permanent; recommends any necessary changes based on pilot project.	Yes. Recommended changes implemented and policy adopted.	Positive evaluation of Tribal certification control & improved government-to-government relationships between State & Tribes.
Policy # 2 NWT	Annual government-run evaluation: describes progress in several social sectors; provides descriptive monitoring of language use.	In the past, evaluation has resulted in direct use (2003 amendments). On-going use may occur in different service sectors, but is not mandated.	Regular publication of descriptive data helps to show progress (or lack thereof) & show that improvement in this area is a government priority.
Policy #3 Council of Europe	Triennial external evaluations: assesses whether states are implementing policies as agreed; recommends ways to improve implementation.	Variable; dependent on member state's willingness to adopt recommendations.	Regular evaluation & publication of results encourages states' accountability if they do not immediately adopt recommendations.

Discussion of Best Practices for Evaluating ML Policies

What have these cases shown about evaluating ML policy? What evaluation approaches are likely to be effective in improving ML policies? In comparing the impacts of these evaluation methods, the frequency and the author(s) of the evaluation appear to influence the potential positive impact of these ML policy evaluations. In the first case study of Washington State's pilot policy, the authors of the evaluation were both community stakeholders and government authorities, and their recommendations were adopted directly. This complies with several of the factors outlined by Leviton and Hughes (1981) relating to whether an evaluation is likely to be successfully utilized. Three of the factors demonstrate the importance of stakeholder (or *user*) participation (communication between researchers and users, information processing by users, and user involvement or

advocacy). A fourth factor, *relevancy*, was present due to the need for the evaluation in order to conclude the pilot program, either positively or negatively. In relation to the frequency of evaluation, however, this one-shot evaluation approach cannot result in any future improvements to the policy.

In comparison, case studies 2 (Northwest Territories) and 3 (Council of Europe) show how regular evaluation mechanisms have the potential to promote on-going improvements, although whether or not recommended improvements are actually used is an additional issue. A lack of participation by policy users in evaluation of both the Northwest Territories policy and the Council of Europe policy may contribute to weakening Leviton and Hughes' (1981) fifth factor, the plausibility of research results. The work by government officials and external experts to assess policy success is useful and necessary, especially where external evaluators or explicit measures are present to increase accountability. However, the greater the distance between the policy evaluators and the individuals, organizations, and communities that the policy impacts, the greater the difficulty in producing results that will be relevant and acceptable to all stakeholders. Considering issues that are prevalent in ML contexts, such as conflict over dialect use, rapid shifts in population, and diverse goals within multilingual and multicultural communities, evaluators who are not familiar with the contexts of policy use, or do not invite the perspectives of users, will be hard pressed to make relevant and credible recommendations. Clear channels of consultation and participation would enhance the quality of these ML policy evaluations. Disseminating results through different channels, and in a variety of languages, in order to make them maximally accessible is also important in order for both direct and conceptual use to occur.

Reflecting back on the methods of evaluating ML policies that have appeared in scholarly literature, discussed above, these case studies lend support to the argument that careful consideration of social ecology is necessary, both at the government level and the local level. As financial considerations are rarely absent from governmental social ecologies and decision-making processes, it is likely that inclusion of cost-benefit analyses, alongside other forms of evaluation, may compel authorities to take action on results, although these case studies do not include examples of cost-benefit analysis. The assessment of language competence is notably absent from all cases as well. If assessment measures based on the education goals of stakeholders (from advanced literacy skills to basic awareness) were developed, this would be another potentially beneficial approach to measuring quality in these policy contexts. In addition to the potential utility of economic and proficiency analyses, quasi-experimental models such as pre- and post-test designs and matching comparisons would be feasible within these ML policy contexts, and might provide compelling evidence for uptake of recommendations. Regardless of the evaluation method, the factors raised by the case studies in this paper—namely the frequency and author(s) of an evaluation—may be additionally useful in enhancing an evaluation's impact on eventual policy improvement. In conclusion, as ML policies increase in number, the way towards improving them is likely to include repeated, methodologically rigorous and participatory evaluation.

Haley De Korne studies Educational Linguistics at the University of Pennsylvania. She is engaged in research and programs that promote multilingual education policy and practice.

References

Bratt Paulston, C., & Heidemann, K. (2006). Language policies and the education of linguistic minorities. In T. Ricento (Ed.), *Language policy: Theory and method* (pp. 292-310). Oxford, UK: Blackwell.

Byram, M. (2008). Developing language education policy in Europe—and searching for theory. *Language, People, Numbers: Corpus linguistics and society, 64*(1), 85-97.

Canagarajah, S. (2005). *Reclaiming the local in language policy and practice.* Mahwah, NJ: Lawrence Erlbaum.

Canagarajah, S. (2006). Ethnographic methods in language policy. In T. Ricento (Ed.), *Language policy: Theory and method* (pp. 153-169). Oxford, UK: Blackwell.

Cooper, R. (1989). *Language planning and social change.* Cambridge, UK: Cambridge University Press.

Council of Europe, Secretary General. (December, 2009). Application of the European Charter for Regional or Minority Languages: Biennial report by the Secretary General to the Parliamentary Assembly. Retrieved December 15, 2010 from http://assembly.coe.int/Mainf.asp?link=/Documents/WorkingDocs/Doc10/EDOC12300.htm

De Korne, H. (2010). Indigenous language education policy: Supporting community-controlled immersion in Canada and the US. *Language Policy 9*(2), 115-141.

European Charter for Regional or Minority Languages. (1998). Strasbourg: The Council of Europe. Retrieved February 9, 2010 from http://conventions.coe.int/Treaty/EN/Treaties/HTML/148.htm

Fishman, J. (1980). The sociology of bilingual education. Lectures at the University of New Mexico, Albuquerque. Cited in N. Hornberger (1988). *Bilingual education and language maintenance.* Providence, RI: Foris Publications.

García, O. (2009). *Bilingual education in the 21st century: A global perspective.* Chichester, UK: Wiley-Blackwell.

Grin, F. (2003). *Language policy evaluation and the European Charter for Regional or Minority Languages.* Hampshire, UK: Palgrave Macmillan.

Grin, F., & Vaillancourt, F. (1999). *The cost-effectiveness evaluation of minority language policies: Case studies on Wales, Ireland and the Basque Country.* Flensburg, Germany: European Centre for Minority Issues (Monograph #2).

Hatry, H., & Newcomer, K. (2004). Pitfalls of evaluation. In J. Wholey, H. Hatry, & K. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 547-570). San Francisco, CA: Jossey Bass/Wiley.

Haugen, E. (1983). The implementation of corpus planning: Theory and practice. In J. Cobarubius (Ed.), *Progress in language planning* (pp. 269-290). Berlin,

Germany: Mouton.

Hinton, L., & Hale, K. (2001). *The green book of language revitalization in practice*. San Diego, CA: Academic Press.

Hornberger, N. (1988). *Bilingual education and language maintenance.* Providence, RI: Foris Publications.

Hornberger, N. (2005). Opening and filling up implementational and ideological spaces in heritage language education. *The Modern Language Journal, 89*(4), 605-609.

Hornberger, N., & Hult, F. (2008). Ecological language education policy. In B. Spolsky & F. Hult (Eds.), *The handbook of educational linguistics.* Oxford, UK: Blackwell.

Johnson, D. C. (2009). Ethnography of language policy. *Language Policy, 8*(2), 139-159.

Leeuw, F., & Vaessen, J. (2010). Introduction: Mind the gap. In F. Leeuw & J. Vaessen (Eds.), *Mind the gap: Perspectives on policy evaluation and the social sciences.* London, UK: Transaction Publishers.

Leviton, L., & Hughes, E. (1981). Research on the utilization of evaluations: A review and synthesis. *Evaluation Review, 5*(4), 525-548.

Leviton, L., & Boruch, R. (1983). Contributions of evaluations to educational programs. *Evaluation Review 7*(5), 563-599.

Menken, K., & García, O. (2010). *Negotiating language policies in schools: Educators as policy makers*. New York, NY: Routledge.

Northwest Territories. (2003). One land many voices: Final report of the special committee on the review of the official languages act. Retrieved December 10, 2010 from http://www.assembly.gov.nt.ca/_live/documents/content/TheFinalReport.pdf

Northwest Territories. (2009). Final report on the review of the Official Languages Act, 2008-2009. Retrieved December 1, 2010 from http://www.assembly.gov.nt.ca/_live/documents/content/09-11-3%20Annual%20Report%20on%20Official%20Languages%202008-2009.pdf

Northwest Territories, Statutes of the Official Languages Act, chapter 23. Retrieved March 5, 2009 from http://www.justice.gov.nt.ca/Legislation.

Pawson, R. (2006). *Evidence-based policy: A realist perspective.* London, UK: Sage.

Rose, R. (1993). *Lesson-drawing in public policy.* Chatham, NJ: Chatham House Publishers.

Rossi, P., Freeman, H., & Lipsey, M. (1999). *Evaluation: A systematic approach.* London, UK: Sage Publications.

Rubin, J. (1971). Evaluation and language planning. In J. Rubin & B. Jernudd, (Eds.), *Can language be planned? Sociolinguistic theory and practice for developing nations*. Honolulu, HI: University Press of Hawai'i.

Ruiz, R. (1984). Orientations in language planning. *NABE Journal, 8*(2), 15-34.

Skilton-Sylvester, E. (2003). Legal discourse and decisions, teacher policymaking and the multilingual classroom: Constraining and supporting Khmer/English biliteracy in the United States. In A. Creese, & P. Martin (Eds.), *Multilingual classroom ecologies: Inter-relationships, interactions, and ideologies* (pp. 8-24). Clevedon, UK: Multilingual Matters.

Spolsky, B. (2004). *Language policy*. Cambridge, UK: Cambridge University Press.

Tollefson, J. (1991). *Planning language, planning inequality*. London, UK: Longman.

UNESCO. (2010). Atlas of the world's languages in danger. Retrieved March 8, 2011 from http://www.unesco.org/culture/languages-atlas/en/atlasmap. html

Washington Professional Educator Standards Board and First Peoples' Language/ Culture Certificate Committee. (2007). End of pilot analysis and recommendations for program future: First People's language/ culture certification program. Retrieved April 20, 2009 from http://www.pesb. wa.gov/FirstPeople/default.htm

Washington State First Peoples' language, culture and oral tribal traditions teacher certification program. (2007). Revised Code of Washington, 28A.410.045. Retrieved March 13, 2009 from http://apps.leg.wa.gov/RCW/default. aspx?cite=28A.410.045

Weiss, C. (1980). Knowledge creep and decision accretion. *Knowledge: Creation, Diffusion, Utilization, 1*(3), 381-404.