Developing a language policy for research in LMICs

Learning Brief No. 2

Katrina Barnes, Annette Zhao, Noor Ullah and Gentille Gasanabandi

Date of publication
April 2023

Author contacts
k.barnes@jigsawconsult.com
a.zhao@jigsawconsult.com
n.ullah@jigsawconsult.com
g.gasanabandi@jigsawconsult.com

Suggested citation
Barnes, K. et al. (2023). *Developing a language policy for research in LMICs*. (Learning Brief No. 2). Jigsaw.

About the Jigsaw Learning Brief Series
The Jigsaw Learning Brief Series provides an open-access contribution to building evidence for education. Each brief focuses on a different issue in education and research in low- and middle-income countries, sharing insight and thought leadership to help shape the sector.

Key messages

Research organisations should:

- develop a language policy that considers the linguistic backgrounds of participants, partners and employees, and ensures that language and multilingualism are considered at both a practical and higher level in research projects;
- budget for employees, translators and enumerators who are fluent in the languages and contexts of the research;
- embrace and value linguistic diversity and ability in research, and integrate multilingualism into every step of the research process to ensure ethical and effective research; recognise that English may not be the preferred language of communication for all participants and partners;
- look to organisations in the fields of linguistics and education for guidance on language policies, as there are few established policies within the research sector.

Introduction

This learning brief is aimed at any research organisations who work with users of languages other than English, or who may do so in the future. It begins by explaining the rationale behind Jigsaw's current language policy. It then considers how working with multiple languages is being considered (or not considered) by other international organisations and extracts potential learnings from these examples. Finally, a toolkit for developing a language policy is offered. This section guides the reader through the different stages common to most research projects, highlighting which linguistic issues should be addressed at each stage and providing examples of how Jigsaw has approached these in past projects.

Why do we need a language policy?

Like many other research organisations, Jigsaw works with different stakeholders including funders, implementation partners, in-country enumerators, and participating communities and individuals in different Low- and Middle-Income Countries (LMICs). Unsurprisingly, in addition to many other localised and contextual factors, language is also a major factor that needs to be carefully considered and managed throughout a research project. In the following discussion, we focus on 'language(s)' in the sense of distinct communication systems, as well as their associated regional and social varieties, as opposed to 'language' in the sense of word and style choices within a given language (including issues around register, taboos, jargon, etc.).

It is often assumed, especially by research organisations based in the Global North, that English will be the main (and sometimes only) language used for data collection, analysis, and dissemination. However, English is in fact not the preferred language of communication for many of the groups with whom Jigsaw works, with some participants and partners having low to no proficiency in English. Participating individuals and communities often exist in complex multilingual contexts, moving between using colonial languages (e.g. English, French, Spanish, Portuguese) in some settings and less documented languages (e.g. Kikuyu in Kenya, or the various Akan languages in Ghana) in others, on a daily basis. Others may operate in local languages only, with no access to colonial languages. This multilingual reality challenges our monolingual way of working: as a first practical step, we need to budget and employ translators and enumerators who are fluent in these languages and contexts. More importantly, without integrating and embracing multilingualism in every step of our research, ethical and effective research in such contexts is not possible.

This reality necessitates a language policy on an organisational level. An established language policy ensures that language and multilingualism are considered at both a practical level (e.g. hiring multilingual enumerators) and a higher level (e.g. embracing and valuing linguistic diversity and ability in our work).

Current status in research organisations

The need for language policies or, at least, significant consideration of language use

when working with participants with diverse linguistic backgrounds has been highlighted previously in academic literature, especially in the fields of linguistics and education (Ganassin & Holmes 2013, Schembri & Jašić 2022). To date, however, very few organisations in the field of international education have targeted policies governing the use of languages, especially among research organisations and those that work with linguistically diverse populations.

In a broader context, there are national and international organisations, often those focusing on languages (assessment, education, translation, and interpreting), that have relevant official statements on the use, accommodation and/or promotion of minoritised languages and multilingualism. However, even extremely well-known international organisations, or those with established practices of working with professional linguists, rarely articulate their language policies publicly, and those that do go into limited detail. A 2011 report from the United Nations stated that 'few organisations of the United Nations system have a formal policy on multilingualism, although the use of different languages in matters related to documentation, meetings and external communications is a general and factual reality.' The ICRC, which works regularly with professional interpreters in conflict zones, gives almost no detail at all on its website regarding its use of languages. An exception to this trend is the Association of Two-Way & Dual Language Education (ATDLE), who clearly stress the importance of bilingual education and dual language education within their key values.

The lack of existing language policies within the research context is mirrored by the lack of relevant guidelines available for research organisations hoping to establish such a policy. As pointed out by Davies & Elderfield (2022), who advocate the inclusion of multilingualism in monitoring, evaluation, accountability and learning (MEAL) activities, current practices discourage full participation from speakers of minoritised languages and ultimately compromise the quality of the research.

A language policy development toolkit

Proposal stage

Typically, this stage involves responding to a 'request for proposals' or a 'terms of reference' (ToR) released by an organisation that requires research, evaluation or strategy work to be done. In this response, the research organisation is expected to provide details regarding how its team would carry out the work required, and at what cost.

The successful navigation and incorporation of multiple languages, including non-verbal languages such as sign languages and braille, depends heavily on appropriate steps being taken at the proposal stage. This includes key information about which languages are used by all parties involved in the research: which languages (and language varieties) are used by the client organisation in their internal interactions; and which are used by participants in their homes, communities, and for official communications. Translators Without Borders (now CLEAR Global)'s language questions tool provides a useful template for determining this information. It is also

important to establish which additional language(s) research stakeholders are able and willing to use when participating in the research or evaluation process, and their proficiency levels in these additional languages (including details of how these levels differ between spoken and written forms). Finally, the languages of project documentation should be established. These details should be located within the ToR, or if they are not present, then project staff should be contacted for clarification.

This level of knowledge about the project's linguistic background is important for two reasons: it helps the proposal team to design an appropriate methodology and workplan sensitive to the local linguistic and cultural context; and it is crucial for quoting an ample budget. As an example, low reading and writing proficiency levels, or the fact that some communities may use one language for written communications and another when speaking, may contribute to decisions to conduct interviews rather than surveys. Furthermore, this knowledge enables the research organisation to determine which linguistic skills will be required to ensure that all stakeholders' voices are heard. In addressing language-related issues, researchers can gather richer and higher quality data that is more representative of the population. In cases where these skills are not present in the team or difficult to source, the research organisation may decide not to proceed with the proposal, concluding that other research teams would be better placed to conduct the study in question. Alternatively, it may just mean that a higher price is quoted so that the budget has sufficient funds for the recruitment of external translators, interpreters, and consultants where required (though this must be done carefully as, in highly linguistically diverse communities, even highly skilled linguists or enumerators from that area may face language barriers). Relatedly, the budget must also be large enough to cover the time needed for language-related activities such as translating survey instruments, transcribing and translating interview responses, and general multilingual communication throughout the project.

It is worth noting that, while it is important to begin this investigative work at the proposal stage primarily for budgetary reasons, many of the finer methodological points can usually be worked out during further scoping activities once the proposal has been accepted, during the inception stage (see below). It should be made clear to the potential client that the budget quoted is indicative, but remains dependent on further scoping.

Whenever possible, the Jigsaw team always ensures that the linguistic expertise needed for a project is considered prior to the signing of the contract. This can be seen directly in the proposal regarding the use of in-country enumerators. During the proposal stage for an evaluation of the iMlango-Transitions programme, an FCDO-funded, Girls' Education Challenge Transition (GEC-T) project in Kenya, the Jigsaw team were able to make contact with a group of potential enumerators from an organisation we previously worked with and put forward a strong case for the inclusion of these enumerators and related budget requirements. Our prior knowledge of this group helped to ensure that they possessed the linguistic knowledge to ensure high quality data.

Similarly, in the proposal for an evaluation of Plan International UK's Supporting Adolescent Girls' Education (SAGE) project, the research team were able to reach out to an organisation who had previously worked on the same project and establish an understanding of their linguistic expertise, before including them in the proposal.

Inception stage

This stage occurs soon after the proposal has been accepted. It builds on the proposal stage in the sense that it is an opportunity to gain a deeper understanding of the project team's aims and priorities for the work requested. This is usually done by holding an inception meeting with the client, reviewing all relevant project documentation, and possibly conducting informal interviews with project staff. This is also the stage at which the study's work plan and methodology is made more concrete.

It is at this point that a deeper understanding of the project's linguistic landscape can be gained. This may involve further research into which languages different stakeholders use and in which situations by accessing census data and language maps (e.g. CLEAR Global's language data). It should also involve an investigation into the cultural significance of key project terms. This is especially important in instances where the research involves vulnerable groups or sensitive topics such as disability or sexual exploitation (see Safeguarding Support Hub 2021). Words that are neutral or generally accepted to be inoffensive in one language may be much stronger or carry more negative connotations when 'directly' translated into another language: referring to someone as 'negro' in many Spanish-speaking communities does not carry the same impact as using the 'direct' equivalent in English, for example (see Adamovsky 2015; Fahrutdinov et al 2017). Similarly, communities who use the same language (but perhaps different varieties of it) may object to the use of certain words for religious reasons, or some terms may be politically loaded in some contexts (think of the difference between terming someone a 'Republican' in the USA, versus using the same term in Northern Ireland). In essence, it is important to remember that language is never neutral, and care must be taken to understand the ways in which key terms will be used and interpreted in the communities involved in the research (see Oxfam's Inclusive Language Guide for an example).

As recommended by the Safeguarding Support Hub, the best way to gain this understanding is by consulting with community members themselves. This could involve presenting community leaders, via implementation staff, a glossary of key terms for their approval. Ideally, they should also be asked to define exactly what they understand by the terms presented to aid later translation. Once this information and approval has been obtained, the research team should arrange for one or several team member(s) with translation experience and working proficiency in the community language(s) to provide a list of translations for these terms in working languages used by the project team (often dominant languages such as English, French or Spanish). This could be done in consultation with project staff who work both in the team's working languages and the community languages

in order to arrive at pairs of terms that are as equivalent as possible.

Finally, the time and language-related duties and the time required to carry these out should be confirmed and agreed with the relevant internal staff or contractors at the inception stage. It is important to get the input of those who will be performing these tasks, as otherwise project managers, who may not speak other languages, may underestimate the time and effort required to do these tasks well.

During the inception stage for the Partnership for Digital Learning and Increased Access (PADILEIA), a project with King's College London investigating higher education access and transitions in Jordan and Lebanon, the research team was able to do vital linguistic planning. Knowledge of the study context within the team made it clear that student participants would need to access surveys and interviews in Levantine Arabic. Building upon this, scoping activities revealed that a consortium member based in Beirut had enough prior experience to translate survey instruments, aided by her deep contextual and project knowledge. In addition, a Jordanian language consultant with a proven track record working on Jigsaw projects was engaged to conduct interviews with students.

Data collection

In this stage the organisation gathers information for decision-making, strategic planning and analysis. Data collection provides the evidence that's needed to answer questions asked in the proposal stage. Data collection happens on numerous levels. The research organisation adopts various techniques and methods to collect data.

When it comes to multilingual projects, team members (or external consultants where necessary) with the relevant linguistic knowledge should be assigned to reviewing all project documentation written in languages other than English. Those team members should then summarise key information in English to enable the rest of the team to access it. This is efficient in terms of budget and a logical initial step to take while working with multilingual projects. Generally, effort should be made to access the documents and papers with the highest degree of usefulness and relevance to the issue in question, regardless of the language in which they are written.

A recent Jigsaw project investigating the potential for EdTech to enhance education in minoritized languages, involved a rigorous review of literature in different languages. To do this, team members who could read the languages in question searched and wrote English summaries of relevant literature. This ensured that the project was able to capture learnings not published in English, and the process enabled all team members to use these sources.

Decisions regarding the design and distribution of surveys for multilingual respondents must be based on key linguistic information gathered at the scoping stage of the project (see 'Proposal Stage'). This information can then be used to determine whether there is a common language that all respondents will be comfortable reading and writing in, to the extent that the survey requires. In cases where literacy levels are low, it may be

necessary to arrange for a scribe to assist participants with recording their responses, or to alter the survey format so that questions are posed verbally. If there is a clear, commonly used first language, then distribution may be done using that single common language. In this scenario, data collection may be greatly facilitated by enumerators with knowledge of both the survey language and other local languages, who can resolve any comprehension issues in real time. It is important to bear in mind that on-the-spot translation can be challenging even those with very high levels of fluency in both languages, especially when technical terms are used. However, a number of practical steps can be taken to prepare enumerators to do this effectively, such as: giving them survey templates or glossaries in advance; facilitating a short tool review to discuss any anticipated comprehension issues; comprehension testing the tool with enumerators themselves to ensure they fully understand all the terms used. This greatly reduces the risk of information loss when the survey is administered, preserving data quality.

If there is no single common language that can be used, all survey instruments should be translated into the preferred language(s) of respondents, even if this means translation into multiple languages. This should ideally be done by individuals with first-language proficiency and experience of translating in those languages, or at the very least quality assured by an individual with first-language proficiency.

When using survey software with multilingual capabilities, automatic translations done using the software should be proofread, ideally by an individual with first-language proficiency in that language. This is because these softwares do not provide 100% percent accuracy, will not factor in contextual details or regional differences, and will not 'know' to use the specific terms established by the glossary creation exercise (see 'Inception Stage'). One possibility is to train staff members to translate surveys using computer-assisted translation (CAT) tools, which can be programmed to 'remember' set glossaries or translate topic-specific phrases in a certain way.

During data collection for a project in Tanzania, the research team utilised data collection software which provided automated translation of data being collected. The team were able to adapt data collection instruments (e.g. observations) to work with the software. This reduced the amount of written notes in Kiswahili the researchers would normally take, thus further reducing the time, effort, and budget needed for translating in the data collection stage (also see Data Analysis below).

All tools should include a short set of questions (perhaps using the aforementioned language questions tool or similar) to determine participants' language(s) and proficiency levels in those languages as part of the general demographic data collection. This constitutes valuable data that can be used for language mapping activities, and can illuminate how language might be influencing respondents' experiences. It could also help to explain any incomplete or ambiguously answered questions within the survey data, especially in cases where respondents' reading and writing

proficiency in that language may have been overestimated. It may be helpful for enumerators (where available) to note their perception of each participants' comprehension and feed this back to the research team, which may help to flag respondents who have not understood and whose responses could therefore compromise data quality. Such information should accompany the data so that researchers can make informed decisions during data cleaning and analysis.

All interview templates should be translated into the language that will be used to conduct the interview, ideally by an individual with proven translation skills and also with first-language proficiency in that language. Where this is not possible, translated transcripts should be quality-assured by someone with first-language proficiency in the target language.

Subject to interviewees' consent, interviews should be recorded. The interviewer(s) should make detailed notes in the interview language, which should be translated into English later, ideally by a first language English speaker within the team. The recording may be used during the translation process to check accuracy and resolve any ambiguities. This process ensures that all interviewees' views are represented accurately, regardless of the language in which the interview is conducted. It also ensures that data gathered from interviews conducted in different languages is of an equally high quality.

The above procedures should form a core component of training given to all external consultants who will be working in languages other than English. External consultants' transcription and translation work should also be spot-checked for quality assurance by members of the research team.

For the Voices of Refugee Youth project, which involved working with young refugees in Pakistan and Rwanda, young refugees were recruited and trained as Youth Researchers. They were able to use their considerable language skills, as well as their contextual knowledge, to ensure that participants were comfortable and able to express themselves fully in their preferred languages. Youth Researchers also provided participants with translated interview sheets to further facilitate engagement with the questions. Later on, Youth Researchers transcribed interviews and translated them into English in collaboration with other members of the research team.

Data analysis

This is the stage at which the data collected is organised and interpreted by the research team. This may involve the use of data analysis software. If the data is quantitative, programs such as R, SPSS or STATA may be used to generate statistics. If it is qualitative, programs such as MaxQDA or NVivo may be used to help researchers draw out important themes. Both forms of data analysis are done with a view to answering the research questions established at the beginning of the engagement.

All transcripts and survey responses should be translated into the research team's working language prior to analysis to allow for all team members to complete analysis tasks regardless of their knowledge of the data collection language. This may be done by either internal team members with the required

linguistic skills and translation experience, or by external consultants. In the first instance, particular attention needs to be paid to the budgeting of researcher time and workload: multilingual researchers can often end up with the majority of the labour-intensive translation tasks, leaving them limited time for higher-level analysis. In these cases, analysis is likely to be led solely by monolingual researchers, leading to a dominance of monolingual perspectives in the actual analysis and reporting and thus reinforcing the monolingual status quo. All translations should be quality-assured by the research team and in consultation with those who collected the data, who are the best place to help resolve ambiguities created through the translation process.

In a project based in West Africa for Plan Ireland, we were able to conduct interviews in English, French and Arabic, thanks to a combination of internal team language skills and the recruitment of external consultants based in each country. External consultants also facilitated the data analysis stage by clarifying ambiguous phrases or sections from transcripts as analysis progressed.

For a project in Tanzania where the team relied heavily on a small group of Kiswahili and English bilingual researchers, transcribing software was budgeted for and used during the initial data analysis stage so that these researchers could focus on analysing already transcribed data. This freed up the limited number of researchers, especially senior ones, to provide insightful analysis beneficial for the project.

Dissemination of findings

This is the stage at which the final deliverables, as agreed by the research team and clients, are produced and shared with stakeholders within and beyond the project. The number and types of deliverables should have been clearly agreed during the proposal and inception stages, and when relevant, multilingual versions of these deliverables should be supplied by the research team to maximise the impact of a research or evaluation project.

The dissemination of findings is of vital importance to all types of projects. Choosing the suitable language(s) to convey the results to key stakeholders ensures that the main messages, implications and recommendations get heard and maximises the impact of a given evaluation, research or strategy project.

The first step of results dissemination is to ensure that the findings are communicated clearly and in a suitable manner in the main working language of the project: most of the time, this language will be one of the more dominant languages (e.g. English). And for many projects, especially strategy projects, this is the sole language of dissemination. In these cases, care must be taken to make sure that the lack of any additional language version does not impede the satisfactory completion of the project. This means making sure that the client and relevant stakeholders are able to understand the findings sufficiently, and other supplements and/or adaptations might be needed to achieve this. For example, it might be worth considering adding infographics with minimum textual information in cases where the working

language does not reflect the linguistic diversity of the client organisation. Adding a summary and/or abstract in a different language might also boost the impact of the final deliverable without requiring much additional resources.

For other projects, producing deliverables in multiple languages may have been a requirement set out in the ToR, or a decision agreed at the contracting stage. It is important to be mindful of the languages and linguistic expertise needed to produce such outputs. Attention needs to be paid to details such as the order of outputs produced: while most projects prioritise English deliverables, it is sometimes helpful to prioritise non-dominant language versions of deliverables to allow in-country implementing partners to transfer learning into practice as quickly as possible. For deliverables that are only available in one minoritised language but not in English or other languages, extra attention is needed so that context-specific learning is captured.

During this step, dissemination to community members such as research participants, especially those from LMICs, needs extra attention. This is because, in most cases, community members are not familiar with the format used in conventional research deliverables (e.g. technical reports to funders and research partners). Producing a version of the report written in the relevant community languages, which clearly highlights the findings most relevant to the community, can significantly increase accessibility.

The platform of dissemination must also be carefully selected in light of the languages involved. For languages widely used in research and evaluation, more platforms are available, especially for non-written outputs (e.g. YouTube for video-sharing). Most software for textual outputs support Latin scripts but might not support non-Latin scripts or languages with no established written convention, thus limiting the means of dissemination. In these cases, it is advisable to seek alternatives popular with the relevant audiences, or used within a similar context.

For research and evaluation projects, Jigsaw often includes a community report in relevant community languages as one of the agreed deliverables. For example, we produced a community report in both English and Arabic for the Partnership for Digital Learning and Increased Access (PADILEIA) project, which investigated higher education access and transitions in Jordan and Lebanon. This 5-page document highlighted the key findings and recommendations of the project in a more accessible format with many infographics, aiming to disseminate the learning to relevant stakeholders at the community level.

General communication during the project

Given the current reliance on dominant research languages such as English, research organisations should actively encourage discussions around languages and multilingualism. It should be established on first communication with project stakeholders which language(s) they are able and prefer to use for communication, including any non-dominant languages which might not have been considered as an option by

some stakeholders. Use of these languages should then be facilitated as far as pragmatically possible. Even if it is ultimately not possible to communicate with stakeholders in non-dominant languages (due to lack of knowledge of that language within the research team, for example), such conversations will at the very least serve to raise awareness about and combat entrenched linguistic power imbalances.

Where a language other than the working language is required, it is a good idea to appoint a research team member with working proficiency[1] in that language to be responsible for general communication (while being mindful of their workload budgeting issue discussed in the Data Analysis section). A multilingual glossary of commonly used phrases in the target language(s) is another option to free up multilingual team members and to enable their monolingual colleagues to engage in basic communication with stakeholders. For translating individual words and short phrases into European languages, tools such as wordreference.com and linguee.eu are recommended. Google Translate may also be used, though it should be borne in mind that its accuracy levels are variable.

Finally, it is advisable for research teams to develop a mini-dictionary/glossary of key phrases in languages used during previous projects, which can be used for basic email communications. This should be used to ensure that communication with project stakeholders is as smooth as possible in cases where research team members who do not speak stakeholder languages need to communicate short, simple messages. For more complicated interactions, communication should be conducted by a team member with proficiency in the stakeholder language.

When working on an evaluation project with Plan Ireland, the Jigsaw team maintained regular communication with in-country staff members in nine countries across two programmes. Due to the limited number of shared languages between the staff from the partner organisations across different countries and the Jigsaw team, we relied heavily on several team members who spoke the relevant languages (mainly Arabic and French) throughout the project. This reliance on certain individuals led to a constant heavy workload for them and as a result, created delays along the way which required even more time from them to communicate with the partners in relevant languages. Learning from this experience, the team has been paying extra attention to the linguistic skills required for general communication and the development of a mini-dictionary/glossary enabled us to better manage similar issues in the future.

[1] 'Working proficiency' is deemed to be equivalent to level C1 on the Common European Framework of Reference for Languages (CEFRL). Speakers with this level are "proficient users of the language, i.e. those able to perform complex tasks related to work and study". Further details are available at https://www.britishcouncil.es/en/english/levels/c1

Conclusion and Recommendations

This learning brief has highlighted the importance of language policies for research organisations working with participants who speak languages other than English, and the complexities that must be considered when working multilingually. It is hoped that the suggestions and examples provided, along with Jigsaw's current language policy, will help other research organisations to achieve high levels of rigour and equity when working on different multilingual projects. The key recommendations stemming from this discussion are as follows:

1. **Incorporate considerations of linguistic expertise in the proposal stage:** Identify all languages used by stakeholders, their proficiency levels, and the languages used in project documentation to design an appropriate methodology and workplan sensitive to local linguistic and cultural context, and quote an ample budget. Reach out to potential enumerators or organisations with prior linguistic expertise for high quality data.

2. **Consult with the community, create tools, and confirm language related duties in the inception stage:** Consult with community members to gain a deeper understanding of the linguistic and cultural landscape of the project. Create a glossary of key terms with community approval, and work with translators to arrive at pairs of terms that are as equivalent as possible. Confirm language-related duties and time required with relevant internal staff or contractors.

3. **Use reliable linguistic information to ensure high quality, multilingual data collection:** Use information gathered during scoping to either find a common language or translate the survey or interview questions into the preferred language(s) of respondents. Engage enumerators or translators with first-language proficiency to ensure accurate data collection. Translate all interview templates and notes and record all interviews for quality assurance purposes.

4. **Conduct rigorous quality assurance during analysis:** Utilise transcribing software during the initial data analysis stage for projects with limited resources or language barriers. Ensure that translations are quality-assured by the research team and in consultation with those who collected the data to resolve any ambiguities created through the translation process.

5. **Maximise impact: disseminate research findings in multiple languages:** Consider producing deliverables in multiple languages to maximise the impact of research or evaluation projects. Prioritise non-dominant language versions of deliverables to transfer learning into practice as quickly as possible, and produce community reports in relevant community languages to increase accessibility. Carefully select the platform of dissemination in light of the languages involved.

As the term ‘learning brief’ suggests, we write this piece with the aim of collaborative learning. We encourage all research organisations to remain open to learning from other organisations and regularly revise and adapt their language policies accordingly. In this spirit, we welcome suggestions for additions and improvements: please contact k.barnes@jigsawconsult.com with your thoughts.