Asia-Pacific Multilingual Education Working Group Regional Webinar #4
Student Assessment and Multilingual Education: Designing Instruments, Gathering Data
3 December 2021, 8:30 – 10:00 a.m. (Bangkok time; GMT+7)

Compiled Q&A with Synthesized Responses

These responses are drafted based on the responses from panel speakers. Please reach out if you are interested in further exploring research or case studies mentioned in the responses.

1. What does evidence say about the effective length in use of L1 before transitioning to L2, L3 as media of instruction?

Current research suggests never working toward a “transition”. The learners’ languages and literacies will always support each other rather than being replaced with new ones. Effective transfer of literacy skills and learning relies on the development of L1 alongside additional languages.

As for establishing a minimum time on L1 development, one example is Kathleen Heugh's research that low-resourced classrooms require even more time than the 4-7 years suggested by Jim Cummins’ research. Quality of L1 and additional language instruction is the key instead of time. Attaining sufficient mastery of L2 to learn new content via the L2 is also critical rather than focusing on the number of years receiving instruction in the language.

Another example of quality vs. time is Thomas and Collier's (1997, 2002) research, based on a massive data set including 42,000 non-English speaking students. The finding shows that students are most likely to achieve educational parity with English-speaking children after at least six years of instructional support in their L1. At the same time, the two most successful instructional models researched by Thomas and Collier were when children received instruction via two languages—L1 and L2 from grade 1 to grade 6. Hence, the key to success was not when the L2 was introduced into the classroom but how the L2 was introduced into the instructional mix in the classroom. In the typical case, children were first taught new content in their L1 and subsequently retaught the same lesson in their L2. Because the lesson's content had already been taught and learned, children could more easily focus on the L2 component of the lesson, learning vocabulary, attending to grammar, and noticing how ideas were expressed and encoded in the L2.

There are four critical conditions that support students’ successful multilingual development according to Thomas and Collier and now others.
• First, the school and learning environment must reflect positively on the value and role of both languages.
• Second, there must be a supportive environment for the L2 outside of the classroom so that children get additional exposure to and engagement with the L2 from family, peers, the community, the media, etc. Without environmental support, developing an adequate foundation in the L2 is more difficult.
• Third, the teachers must be highly bilingual in both L1 and L2 to facilitate children’s learning of the L2. This point allows students greater exposure to languages, even if it happens informally in the classrooms.
• Lastly, a gradual introduction into the L2 from early on reduces the shock of being taught solely in L2 before having adequate knowledge and skill in L2.

The research from a large-scale study in Asia demonstrates that children need to have at least a level 3 proficiency out of 5 in L2 to experience general educational success when being instructed solely via L2 or L3. Level 3 is essentially the level of proficiency characterized by Cummins (various dates) as the CALP level of proficiency. Cummins and others suggest that it takes 5-7 years in a school context to reach this level of proficiency.

2. **Is it necessary to assess students’ performance in content areas in a separate language (whether L1 or L2) or while students use all of their linguistic repertoire (both L1 and L2) in case students are still developing each of their languages?**

It is always possible to assess languages separately, but bilingual/multilingual assessments could be helpful. In South Africa, bilingual side-by-side science tests allowed learners to answer any question on either side. The goal was to demonstrate understanding of the content, not testing a language level. Also, researchers and practitioners are currently working on multilingual language assessments. Like this assessment, many scholars are presently considering “translanguaging” or allowing learners to write assessments in a mixture of languages, for assessing contents like sciences, and this would require teachers to be proficient in both/all languages.

Importantly, the necessity and validity depend on the purpose of the assessment. In the case of formal or standardized testing, if the primary goal is to compare two models of instruction—L1 medium vs. L2 medium for members of the same ethnolinguistic community, then the assessment should be done in the language of instruction. Assessment done in this circumstance focuses on the relative effectiveness of the two models, so it should proceed in a manner that reflects how the model is functioning. If/when the purpose of the assessment is to measure the amount of content that students are mastering, then the answer to this question is less obvious and may well vary depending on the age or grade level being assessed. For example, while learning to read is a very language-specific task, especially during the learning phase, this is not necessarily the case concerning math. In terms of classroom/individual assessment for students’ learning progress, the use of the L1 may well be more productive for assessing their experience than using L2.

Research conducted by Abedi investigates the test performance of second language speaking children in the United States. When the testing dealt with more language-dependent contents
such as reading comprehension, science, and social studies instead of math, there were significant deficits in students’ test performance in their second language even though they learned in the language. If limited knowledge of a language serves as a barrier to learning, measured deficits in learning may be more likely due to actual deficits in learning rather than the possible constraints of the language of the assessment.

3. **Can transferring skills from L1 to L2 and L3 be measured?**

Teaching for transfer means developing comparison/contrast charts and making the similarities and differences between languages explicit and clear. Assessments of these comparisons/contrasts would thus be possible and quite a good idea to see how far learners have come in transferring skills between their languages.

Some skills and knowledge are language-independent, and the others are language-dependent. Transferrable skills are language-independent that can be applied in various languages. Language-independent skills embedded in the reading process, which, once mastered, can be used in other languages without being “relearned.” For example, mastery of the alphabetic principle of how words are represented symbolically with combinations of symbols or letters is not a skill or piece of knowledge that needs to be relearned every time one learns a new language. While it is a commonplace that a given language contains vocabulary encodings not found in any other language, linguists recognize that few languages have structural features not seen in any other language. Hence, the more one learns about languages in the comparative sense, the more one recognizes the level of commonality across languages.

On the other hand, learning to read is very language-dependent, which is a compelling reason why children learn to read in a familiar language more readily than in a lesser-known or unknown language. In this sense, testing one’s reading ability in English, for example, makes little sense when the language of the assessment is Swahili. If the one being tested does not know what the task is for which a response is expected, any response on that person’s part is little more than a random guess as to what the real task was.

Although language is a vital mediating factor in how knowledge and many skills are passed from one person to another, a large portion of the knowledge and skills learned in school are not unique to a given language (language-independent). The skill is not “transferred” from one language to another. Rather, a given person, a child or student, experiences development in their language skill, allowing that person to describe or explain the skill in more than one language. Such a skill can be assessed in more than one language. However, when doing so, it must be remembered that the purpose of the assessment should focus on measuring the growth of language mastery, not the existence or manifestation of the skill.

4. **How should MLE be designed for bilingually homogenous communities vs. bilingually diverse communities?**

The basic principles are the same in both situations—developing strong skills in the L1 while exposing learners to one or more additional languages and constantly checking for understanding in their strongest languages.
Here are some suggested meanings of bilingually homogenous community and bilingually diverse community with some community-specific suggestions:

**“Bilingually homogenous community”**: a high degree of linguistic homogeneity in a minority or lesser-known language with that community being also embedded in a sociolinguistic content. Language of wider communication or official language is prominent in education, the market, the media, and governmental functions. The design and implementation of an MLE program in this community are much more straightforward. An educational foundation is built for all children, primarily using the L1 common to virtually all the children.

**“Bilingually diverse communities”**: Multiple minority languages exist side-by-side even in the same physical community. A more established or official language is generally present in education, business, media, and government.

To design an MLE program for these bilingually diverse communities, the following steps might be helpful in general:

- Find out what language the children use in the street when they play with their friends and peers. In many cases, a particular language tends to emerge as the “lingua franca” of the community or region.
- Explore the possibility of using this lingua franca as the language of instruction in the schools, which is a conceptually ideal solution. This language will tend to be the most familiar or common language of the most significant number of children. The most typical solution, in such cases, is to use the national or official language as the medium of instruction because knowledge of the official language will tend to be considerably higher in regions of intensive mixing of languages.

Once the main languages are identified, there are three possible scenarios:

- The community/school decides on one primary language to work on/in—bilingual program and figures out ways to include other languages.
- The community/school decides on two main languages and a dominant language—a trilingual program.
- [less attractive] The school still functions with only a dominant language, but all languages are welcome in the classroom, and there are classroom aides (parents) or other ways to support learning.

In an urban context in which immigrants tend to settle into well-defined concentrated areas, there is a greater possibility that an L1 first strategy is viable. When there are no such concentrated areas, using the most common or most widely spoken language at the children's level offers the most significant potential for instruction in a familiar language.

Stephen Walter’s research findings offer a country example where children’s L1 is the medium of instruction as well as a widely spoken language. He analyzes that there are two settings where linguistic mixing takes place: (1) the cities (2) the areas where two languages came into regular contact. The country in focus addressed this phenomenon by encouraging parents to send their children to the local school (or the closest school) using the L1 of the child as the language of
instruction. Parents did not always follow (or might not have been able to follow) this guidance. For the vast majority, however, this strategy proved generally workable.

5. **As a country like Myanmar continues to have education in an emergency stage (conflict situation) in supporting ethnic students' learning, how can education stakeholders adjust assessment/ data validation strategies and ensure that these situations increase learning opportunities while reducing inequalities?**

It is crucial to consider the purpose of comparing assessment results. The main idea of assessment should be to provide teachers, school directors, and supervisors with diagnostic information about learning to improve teaching and learning. Nevertheless, for the assessment, there are no one-size-fits-all methods. The key is to start with where each ethnolinguistic group is in their educational development. Some non-dominant languages in Myanmar have L1 as the medium of instruction through higher education; others have L1 only a few years. Well-done MLE classrooms and programs tend to have a strong positive impact on learning outcomes which, in turn, tends to have a positive impact on educators looking for results and on local communities looking to see their children educationally empowered. If L1 education is possible and in place, it is recommended to work hard to make it as effective as possible and find ways to publicize the results.

6. **Does the fact of wide linguistic diversity suggest that assessment of foundational skills in LOI is more useful in upper primary rather than lower primary. e.g., using the Teaching at the Right Level (TaRL) approach developed by Pratham in India?**

The key is not the timing of the assessment but the learner’s learning progress and understanding of the learning contents. Assessing in a language of instruction that children do not speak or understand will result in a lot of zero scores, which removes the opportunity to diagnose needs. It is essential to consider whether the assessment results indicate a lack of understanding of the content, lack of understanding of the language, or ability to produce enough language to demonstrate understanding.

In Niger, West Africa, Mart Hovens (2002) tested learners from the same L1 backgrounds in two types of schools: bilingual (L1/L2) and non-bilingual (L2 only). The learners were tested in both languages through oral testing. The study produced the following results:

- Best: Those learning in L1/L2, tested in L1
- Next: Those learning in L1/L2, tested in L2
- Next: Those learning in L2 only, tested in L1
- LAST: Those learning in L2 only, tested in L2

The results show that even students who had not been taught in their L1 still were able to express knowledge in their L1 better than in their L2.

In some form of standardized or competitive testing assessment, it is critical to teach teachers to do regular and less formal assessments at the classroom level from the beginning of the educational process. While it may not be common in some contexts, individual-level and classroom-level assessment should be a part of standard educational practice. The absence of
assessment allows teachers to go ahead at a pace established by curriculum writers rather than at a pace dictated by children’s learning patterns. Conducting assessments can prevent children from getting left behind from classroom instruction.

7. How about contexts where there are three or more languages as LOI in the education system? Is it feasible to develop both orality and literacy in 3+ languages within K-6?

It is feasible to develop both orality and literacy in 3+ languages within K-6. However, it is recommended that all three languages need to stay in the education system, and a robust system of comparison/contrast (teaching for transfer) would need to be in place. Carol Benson and colleagues wrote about the Rajbanshi-Nepali-English trilingual program in southeastern Nepal and a UNICEF report by Dhir Jhingran discusses this topic.

Orality and literacy in a language may be understood as follows:

- **Orality in a language** – sufficient knowledge and skill in a language such that a person can function reasonably competently in any everyday social context.

- **Literacy in a language** – sufficient mastery of the symbolic system of a language such that a person can read with reasonable fluency (about 60 or more words per minute), understand the content of what is being read (with 60-90 percent accuracy), and ability to write reasonably accurately so that it can be read at a rate of 10-25 words per minute.

By this standard, developing both orality and literality in 3 languages is not feasible. Judging by many developing countries, only a small/modest percentage of children would develop passable orality and literacy in even one language if that language is not their first language. Bright or gifted children with very capable teachers may develop some skill in two languages, assuming neither is their first language and that one or both are widely spoken outside of the school. The notion that loading more and more into a curriculum to produce highly educated children most often leads to the opposite—content overload leading to attrition, frustration, and failure. For average children, it is better to learn a modest curriculum well than an ambitious curriculum minimally or not at all. In practice, there are few if any schools in developing countries that have successfully developed reasonably good orality and literacy skills in most or all of their students in three languages during K-6.

8. How can children who use non-oral linguistic modalities such as signing languages be integrated into mainstream classrooms and assessed for learning progress?

While this webinar is focused on assessment in MLE, the questions about how to include deaf children are essential to consider if the assessment is to include everyone. MLE principles should have clear applications to non-oral linguistic modalities. such as the following:

- Teachers need to be proficient in both/all languages.
- Both/all languages should be used for instruction, and the L1 should be the basis whenever possible.
- The transfer should be promoted by exploring similarities/differences.
9. What are some suggestions for MLE advocacy to persuade schools to adopt a mother-tongue based approach?

There may be persuasive arguments in assessing learners in MLE and non-MLE programs because MLE student assessment results will be better in the home language and additional languages such as L1, L2, and beyond. Carol Benson’s slides are available on the event webpage for explaining the importance/efficiency of MLE to others.

One of the most compelling arguments for MLE advocacy is a successful demonstration project. This is exemplified by a common expression in English, “Nothing succeeds like success.” Observing a program or activity that works is a powerful argument for that program or strategy. However, the demo needs to be larger than one school since the performance of just one school can be deceiving, especially if extraordinary resources were focused on just that one school.

In terms of parents’ perceptions, presenting the validity of MLE with students’ improved English skills may be effective. When Stephen Walter conducted an MLE research project in Cameroon, one of the persistent responses from parents was wanting their children to learn English, the medium of instruction in the local schools, because they understood English to be the pathway to success for their children. However, when parent respondents saw their children scoring higher on assessments in English in an L1 school, they quickly began to push for a wider implementation of an L1 instructional model because they saw it leading to better English than the traditional English-medium instructional model.

Another compelling argument for MLE advocacy is published research. High-level educational stakeholders at local, state, and national levels depend on published research to inform evidence-based policy planning adapted to their own context.

10. On the use of writing for language assessment, how do we avoid 'subjective scoring'?

Teachers are the best people to assess, but they may need help developing the assessments and the criteria for scoring them. Criteria will deal with subjectivity to some degree, and the rest is just to be expected in humans. The question of subjectivity comes from a positivistic position that presumes that testing has no human elements.

One way to minimize subjective scoring is the use of a rubric. A rubric is a detailed evaluation metric worked out before scoring begins. The rubric guides the scorer(s) on how to evaluate the many features of a response that has no obvious right or wrong answer. Below is an example of a rubric for evaluating a writing task.

The use of the rubric:
- First, dimensions are identified as reflecting as many aspects of the writing task as one wishes to evaluate
- Second, a 3-5-point scale is constructed for each dimension to evaluate each response on that dimension.
When the scoring for a single response is completed, the scores for each dimension are entered into the last column and then added up to give an overall score. Suppose the overall assessment is qualitative or one wishes to have a single qualitative measure of performance on the task. In that case, one can convert the final score to a qualitative measure by equating a range with a word or phrase indicating the overall performance quality on the task. The following is just one of a number of possibilities: 0-4 = Very weak, 5-8 = Below average, 9-12 = Acceptable, 13-16 = Excellent.

More detailed or less detailed rubrics can be written or developed depending on the level of guidance needed for the scorers and the nature of the task being evaluated.

<table>
<thead>
<tr>
<th>Feature</th>
<th>No credit (0)</th>
<th>Inadequate (1)</th>
<th>Weak (2)</th>
<th>Acceptable (3)</th>
<th>Very good (4)</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>Nothing written</td>
<td>Only a couple of words added</td>
<td>One completed sentence</td>
<td>At least two sentences</td>
<td>A paragraph or more</td>
<td>2</td>
</tr>
<tr>
<td>Naturalness</td>
<td>Totally mixed up</td>
<td>Decipherable but with many errors</td>
<td>Sort of the way it would be said in English</td>
<td>Reads pretty well but one or two awkward expressions</td>
<td>Pretty much written the way a native speaker would write</td>
<td>1</td>
</tr>
<tr>
<td>Coherence</td>
<td>Unintelligible</td>
<td>Barely intelligible</td>
<td>Discernible thoughts or theme but not well organized</td>
<td>What was written hangs together quite well</td>
<td>Thoughts well organized and expressed</td>
<td>3</td>
</tr>
<tr>
<td>Grammaticality</td>
<td>None</td>
<td>Many errors</td>
<td>Only a few errors</td>
<td>Maybe one or two errors</td>
<td>No more than one error</td>
<td>2</td>
</tr>
</tbody>
</table>

Using assessment results to compare countries may be a misuse of assessment. The purpose of assessment results should be important feedback on teachers’ teaching and work on assessment from the ground up. For example, looking at assessment results showing that children in Bangladesh are decoding 30 words per minute, and children in Papua New Guinea are decoding 31 words per minute, the crucial questions are missing:

- In what language(s)?
- Are these language(s) the ones in which they are learning?
- How are they being taught to read?
- How are teachers trained in reading pedagogies?
- What are the specificities of the script(s) (writing systems) of their language(s)?