The assessment of mathematical knowledge in elementary level dual language programs

To appear in:

Evaluating bilingual education programs: Assessing students' bilingualism, biliteracy and sociocultural competence

Editors: Margarita Machado-Casas, Saúl I. Maldonado and Belinda Bustos Flores

Kip Téllez

University of California, Santa Cruz

Chapter Introduction and Overview

For all their promise, Dual Language Programs (DLP)[1] are a leap of faith for the educators, parents, and students who participate in this pedagogical and social experiment. Of course, all education is experimental—or perhaps should be—but DLP programs push the boundaries of typical school practices. First, routine elementary schools do not attempt to teach an entirely new language to all the students; second, most elementary schools do not explicitly seek out two different groups of students, who often differ by culture and economic background,

[1] The editors of the volume have chosen to use the term Dual Language Programs, instead of others (e.g., two-way immersion, two-way bilingual education). I find the term useful and have followed suit.

and put them into the same classroom, regardless of their existing academic knowledge. Third, most schools are not required to convince parents, administrators, and the local community of their mission and methods; DLP programs are largely optional and must recruit families who willingly and enthusiastically participate.

These additional tasks, regardless of the value of DLPs, take time and energy, both from the general curricula and the educators themselves. I would also add that the stakes are higher for DLP: Poor academic programming may result in students who fail to gain grade level literacy skills in either English or the Language Other Than English (LOTE).[2]

Given the challenges of implementing an effective DLP, effective program evaluation is paramount. Data must inform instructional and curricular decisions regarding literacy growth in both languages (and the relation between the two), progress in oral language development, and student learning in other academic areas. The existing literature on the evaluation of language development in DLPs is established, although we could make the case that better tools and additional evaluation is needed. However, few DLP program evaluations include a deliberate assessment of mathematics instruction in both languages, although several excellent evaluations (e.g., De Jong, 2002) of DLP include mathematics scores and analyses.

With the expansion of DLPs nationwide and the shift to the Common Core State Standards (CCSS), which regards mathematics as a form of literacy, I believe it is time to reassess how mathematics achievement in both languages in DLPs is assessed, analyzed, and used for program improvement.

---

[2] The majority of DLPs offer instruction in English and Spanish, but I prefer to use the more general term LOTE in this case to recognize the many programs teaching English and a language other than Spanish (e.g., Mandarin). However, the assessment examples provided are in Spanish and the programs I highlight are all Spanish/English DLP. However, most of the other information should be relevant for any DLP, regardless of which languages are taught. The chapter is clearly designed to inform DLP programs in the US, but the foundational suggestions will apply to international contexts.

My goal in this chapter is to share with interested educators my own experiences evaluating DLPs, with a specific focus on using mathematics assessments to guide programming. I intend to be practical in my advice, thus avoiding much discussion on the wider issues in the evaluation of language programs. Those interested might wish to read an earlier work in this area (Téllez, Flinspach, & Waxman, 2005). Neither do I have the space to explain many of the technical tools needed by evaluators to test, for instance, statistical significance or effect sizes. Other resources and texts of interest (e.g., Creighton, 2006) offer educators the opportunity to develop their own skills.

My argument for including mathematics assessment in DLP evaluations is three-fold. First, mathematics is language and language is mathematics, and with the new CCSS, the literacy skills needed to understand mathematics has increased. Educators are still orienting themselves to this new way of conceiving mathematics instruction, but the concept of "academic language" is certainly part of these new understandings. Indeed, the definition of academic language itself and how to teach it are still inchoate, but the general view is that learners must develop a discourse of the discipline. For instance, the CCSS suggest that 4th grade students be able to "Construct viable arguments and critique the reasoning of others" (p. 29, CCSS, 2015). Students cannot achieve this goal without the capacity to "speak like a mathematician." DLP educators must therefore consider carefully if mathematics will be taught in English or the LOTE. Regardless of the language of instruction, how will program leaders know if the students are meeting grade level standards? I explore the meaning of academic language more fully in the next section.

Second, because DLP programs have typically been highly focused on literacy development in both languages, mathematics achievement may be lost in the wider language goals. This lack of focus is unacceptable. Whether or not educators agree with the role mathematics achievement plays in educational achievement, students who do not master mathematical content, especially algebra, find their options for postsecondary education curtailed. Some have even called mathematics achievement the new civil rights cause (Moses & Cobb, 2002). Furthermore, Latina/o/x students, who represent the majority of students enrolled in DLP, tend to score lower in mathematics than their white or Asian counterparts (Téllez, Moschkovich, & Civil, 2011). Without careful tracking of their mathematics achievement progress in DLP, these students could fall further behind.

Third, collecting data on mathematics achievement offers teachers another point of data to discuss and learn from. Teachers tend to misunderstand the role, purposes, and analyses of assessments outside of their teacher-created tests, but standardized tests offer much information to consider. And it has been argued that teachers who understand assessment are better able to instruct language learners (Téllez & Mosqueda, 2015). I recognize that additional testing takes time away from instruction, but assessments can be used to thoughtfully inform instruction.

Program evaluation methods have not typically been taught in programs preparing teachers or school administrators, which I suppose is good news for evaluation specialists like me, who are paid to assist schools collect and analyze program data. But the truth is that I would prefer to have school personnel grow their own skills. To this end, I hope the chapter will help DLP educators address the following questions regarding mathematics achievement in DLP:

1. How are our DLP students performing in mathematics, in both English and the LOTE?
2. Does their achievement compare to similar students who are not in a DLP?

3. Are our DLP students making sufficient gains in mathematics?
4. If our DLP teaches mathematics in the LOTE, how can we use mathematics achievement data to inform the decision about when to switch to teaching mathematics in English?
5. What are the advantages and disadvantages of various mathematics assessments (e.g., state mandated, commercially available, district created, teacher-made)?
6. If our DLP teaches mathematics in the LOTE, how much does it help students' literacy development?

In the remainder of this chapter, I hope to convince DLP educators to use thoughtful program evaluation strategies to improve the educative experiences of their learners. As John Dewey pointed out in his groundbreaking work, *Logic: The Theory of Inquiry* (1938), the tools of the scientific methods can be used to judge the value of programs for the betterment of the human condition. As a program designed, at least partially, to lower the social and economic barriers that monoglossic societies routinely erect (del Valle, 2000) and thus improve the overall human experience, as well as enhancing the life chances of the students, the oversight of DLP programs must be informed by careful evaluation and reflection on quality data.

Assessing growth in mathematics in two languages informs growth in bilingualism and biliteracy

The link between mathematical knowledge and linguistic knowledge is strong and growing stronger as a result of contemporary mathematics curricula, in which language is considered an important resource for learning mathematics (Moschkovich, 2012). My intent in this section to offer additional evidence that the assessment of DLP students' mathematical knowledge informs literacy knowledge, recognizing that language development (reading, writing, speaking, listening) tends to be the primary focus of most DLPs.

We can begin with a brief overview of the relation between mathematics and language growth by recognizing that mathematics *is* language; not *a* language, but language itself. Scholars such as Pimm (1987) and others (e.g., Spanos, Rhodes, Dale & Crandall, 1988) have dispelled the view that mathematics is a "language free" discipline. Research by MacGregor and Price (1999) found that a general knowledge of syntax in language is associated with mastering the syntax of algebra. Furthermore, Danesi (2003) has shown that knowledge of metaphor and figurative language is key to understanding and solving "story problems."

The relationship between language and the various disciplines is generally referred to as "academic language" (AL) or the academic "register." Definitions of AL are varied, but a general consensus is emerging, as Snow (2010) suggests. In her view, AL refers "to the form of language expected in contexts such as the exposition of topics in the school curriculum, making arguments, defending propositions, and synthesizing information" (p. 450). AL exists in all academic disciplines, and some intersect through genres and registers. Mastery of the words, symbols, and forms used in a particular specialization (Halliday, 1993), such as the mathematics register, are considered a key element of school success in mathematics competence (Celedón-Pattichis, Musanti & Marshall, 2010).

Teachers and researchers now realize that ordinary words and symbols have significantly added or altered meaning in different contexts of math. For instance, the terms "of" in English and "de" in Spanish are transformed from their common definitions to entirely different meanings in mathematics (e.g. "what is 3/4 *of* 12? / *que es* 3/4 *de* 12?" This simple example offers evidence that mathematics is a language in and of itself, complete with a syntax, punctuation and symbols that serve as “nouns” and “verbs”. Nominalizations are also very common in the mathematics register (e.g., subtract to subtraction). Mathematics AL plays a role in providing students with access to a rigorous mathematics curriculum, but it also may help DLP students become aware of meta-linguistic aspects of mathematical language and thereby enhance literacy skills. The larger point is that mathematics AL must be explicitly taught.

Research has made clear the link between mathematics and literacy achievement (Chen & Chalhoub-Deville, 2016; Han, 2008; U.S. Department of Education, 2008). Indeed, in their comprehensive study of the relationship between mathematics and literacy, Abedi and Lord (2001) argue “that the interaction between language and mathematics achievement is real. This interaction must be a critical consideration in future mathematics assessment research and practice” (p. 232). These studies are, however, correlational and therefore do not necessarily imply that improved mathematics achievement will boost literacy achievement or vice versa, but understanding the exact causal connection would require knowing much more about the brain than our current knowledge allows. Nevertheless, I am convinced that learning mathematics “bootstraps” language learning, especially in the upper grades (Cheng, Li, Kirby, Qiang, & Wade-Woolley, 2010). Mathematics learning adds to a learner’s ability to make logical connections between symbols and other abstractions, which is, more or less, what all AL requires learners to do.

DLP educators are well aware of the claim of cross-linguistic transfer; that is, the literacy skills gained from learning to read in one language transfer to when reading in a second. This simple claim needs specificity, and Cummins (2008) has suggested several categories of linguistic transfer: (a) conceptual elements (e.g., understanding the concept of photosynthesis across languages); (b) metacognitive and metalinguistic strategies (e.g., strategies of visualizing, use of graphic organizers); (c) pragmatic aspects of language use (willingness to take risks in communication through L2); (d) specific linguistic elements (knowledge of the meaning of "photo" in photosynthesis); (e) phonological awareness—the knowledge that words are composed of distinct sounds. I suggest that cross-linguistic transfer relates to literacy development *and* mathematics education. For example, young DLP students must have a deep conceptual understanding of mathematics operations in order for linguistic elements to transfer from one language to the other.

Most of these cross-linguistic transference effects also apply to cross-*content* transfer; that is, content learned in literacy will transfer to learning in mathematics and vice versa. In fact, metacognitive transfer might be more pronounced in cross content transfer. For instance, students who can use graphic organizer to show cause and effect, for example, in a story, will also gain the metacognitive skill to show cause and effect in mathematics (e.g., "if I change the exponent in a function like y=x^2, what happens to the line?"). In the main, if DLP students are not learning deep, conceptual knowledge in mathematics, they are losing ground in their literacy growth, especially as they move up in grade level, when the emphasis on understanding logical inferences in a text increase. Assessing cross-linguistic and cross-content transfer is why DLP programs teaching mathematics in a LOTE must assess mathematics achievement in the language of instruction.

Practical aspects of assessing mathematics in DLP

I hope that the former section has pointed out the potential for mathematics learning to assist in literacy learning, regardless of the language of instruction. Cross-linguistic and cross-content transfer both play an important role. The proper assessment of mathematics achievement in DLP, if nothing else, will likely increase the attention paid to mathematics instruction. A truism in education is that what gets assessed is what gets taught. Leaving aside the negative connotation of the statement, it has been my experience that many DLP educators tend to give too much attention to literacy growth in the language of instruction—especially English—while obsessing over various models (e.g., 90/10, 70/30) when the more important concerns are what is being taught in which language and how.

I will move on now to several important points to consider if DLP leaders wish to better assess their students' mathematical knowledge and how the data might be used to improve their program. I will address four general topics in the remainder of the chapter: (a) the general assessment of mathematics learning, (b) existing resources for evaluating DLP, (c) specific mathematics assessments and their advantages and disadvantages, and (d) examples of three DLP evaluations in mathematics taken from different contexts. I end the chapter with some general advice for DLP evaluations.

The general assessment of mathematics learning. Although a full review of the evaluation of mathematical achievement is beyond our scope in this chapter, a brief overview to provide readers with some context before we begin exploring the assessment of mathematical knowledge in DLP programs. Like other content areas, the assessment of mathematical knowledge has been beset with arguments about our capacity or interest in measuring procedural vs. conceptual knowledge (Crooks & Alibali, 2014). As might be predicted, educators are generally more interested in assessing students' conceptual knowledge, but as with most dualisms, this one is fuzzy. As a shorthand example, the following will suffice: Assessing procedural knowledge using addition might be represented by the following problem: 3+4+5= __, while the assessment of conceptual knowledge using addition can be expressed as an equivalence problem (e.g., 3+4+5 = 3+__). Putting aside the instructional approaches for teaching simple addition vs. equivalence, it is generally easier to measure procedural knowledge, so this is what tends to get measured. Without venturing too deep into psychometric theory, the assessment of procedural knowledge yields greater reliability while measuring conceptual knowledge yields better information about how students will perform as they move up in grade and thus mathematics complexity. Therefore, DLP educators should always choose to measure as much conceptual knowledge as possible, regardless of the language of instruction.

Existing resources for evaluating DLP. The good news is that we have already have a few explicit guidelines for assessing DLP. The most extensive resource, by far, for evaluating DLP was developed by Lindholm-Leary and Hargett (2006). The document was sponsored by the California Department of Education and is available at the Center for Applied Linguistics web site (see URL in References). The authors must be recognized for creating an extraordinarily thorough guide to evaluating DLP, which I have to assume was motivated by the many schools required to provide an evaluation of their DLP to their funding agencies. The requirements for a formal evaluation of a funded project are greater than those for an "internal" evaluation of a DLP for a local audience, and Lindholm-Leary and Hargett do an excellent job of providing the sample documents and templates needed to document the achievement of DLP students. They also provide sample surveys for parents, students, and teachers. Although they do not specifically address the assessment of mathematics in a DLP, the general guidelines they provide would work for mathematics evaluation.

The strength of this toolkit is its comprehensiveness and the “step-by-step” guidance. However, I do not think the data analysis advice is practical. For instance, chapter 9 of the document is a 100-page guide to analyzing data utilizing spreadsheets and specialized statistical software, which most districts do not purchase. If DLP educators were to read this chapter and decide it was worth embarking on such an analysis, I would recommend they enroll in a graduate degree program in quantitative research in education, an endeavor which I endorse without reservation, but which I think would take more time than many DLP educators have to spare. A better option, in my view, is to (a) work with the school district’s research and evaluation staff (if such a staff exist), (b) reach out to researchers from local universities who might be able to assign a graduate student to the project, or (c) hire a professional evaluator (The American Evaluation Association has a searchable list of professional evaluators (https://www.eval.org/p/cm/ld/fid=108). I will have a few more comments to share on this topic at the end of the chapter. Nevertheless, the toolkit is free and clearly worth an examination.

Another good resource, also available from the Center for Applied Linguistics website, is the Spanish-Language Assessments for Dual Language Programs (Sugarman et al., 2007). This useful guide lists approximately 25 assessments (all available in Spanish suitable for DLP evaluations. For our purposes, they list two mathematics assessments in Spanish: the Spanish Assessment of Basic Education Version 2 (now out of print) and SUPERA, which I have included in the list of mathematics assessments below.

Finally, DLP educators can rely on other, more general guides for evaluating educational programs. There are several excellent resources in this category, but I recommend Morrison and Harms (2018). This text offers excellent advice on how to present the results of a program evaluation and is especially helpful if teacher professional development is part of a funded DLP project. I also recommend Creighton's (2006) book which shares techniques for analyzing data that clearly explain the concepts behind comparing scores between two groups. However, as mentioned previously, it is often easier to find an expert to assist with data analyses than to learn these techniques from scratch.

Specific mathematics assessments and their advantages and disadvantages. I imagine this section will be of great interest to DLP educators, but I must first offer several caveats before describing specific assessments. First, the publishers of assessments are constantly changing and often tests are discontinued as publishers try to keep only the revenue generating products in their catalog. Next, assessments go out of print on occasion. As I noted, the Spanish Assessment of Basic Education (SABE/2), which was included in an earlier guide for DLP assessment, is no longer in print, even though the guide was published in 2007. In addition, the names of tests seem to change every few years. Therefore, I make no guarantees that this list will be accurate even one year after the publication of this book. Second, some of these assessments are expensive and require specific qualifications to purchase them, although it likely that at least one professional in the school district, a psychologist with a doctoral degree for instance, can purchase them. Third, some of the assessments are individually administered (i.e., a staff member has to give each student the assessment individually), which makes the assessment very time consuming.

Table 1: List of assessments suitable for evaluating mathematics learning in DLP, including descriptions, benefits and disadvantages of each

Assessment Title or Series / Publisher	Description	Advantages and Disadvantages
State-mandated assessments / Various Publishers	The Every Student Succeeds Act requires each state to assess reading and mathematics at grades 3-8 and 11, so all have developed a valid and reliable assessment. In California, for instance, which is part of the Smarter Balanced Assessment Consortium, the mathematics test in closely aligned to state standards.	Free to schools. Linked to state standards. Offered only in English in some states. Smarter Balanced Assessments are computer adaptive so results should be more reliable than "fixed form" paper and pencil.
Logramos® / Houghton Mifflin Harcourt	Parallels the scope and sequence of the *Iowa Assessments, also published by* Houghton Mifflin Harcourt	In addition to mathematics, it also assesses social studies and science. National norms are provided, so schools can make comparisons to populations.
Aprenda®: La Prueba de Logros en Español Tercera / Pearson	Stanford 10 edition (Pearson) offers a parallel English version.	Very commonly used tests of achievement across a wide range of subjects.
STAR Mathematics / Renaissance	Equated Spanish and English versions.	May not be aligned to state standards. A literacy test in Spanish and English is also available
Batería III lor IV NU Pruebas de aprovechamiento / Riverside	Parallel to Woodcock Johnson III. While both the Batería and Woodcock Johnson III are achievement tests, they are more commonly used to assess students for placement in specialized programs.	Spanish test equated to English versions, which allows direct comparisons between languages and content Tends to be used for diagnostic purposes, which can provide very detailed information on student performance.

SUPERA PLUS/ CTB McGraw-Hill	McGraw-Hill (CAT/6) parallels the TerraNova in English English	Requires purchasing the "PLUS" version to assess mathematics computation "Open-ended" responses used in some of the mathematics items, which allows for some literacy assessment in mathematics.
Locally developed assessments	DLP leaders often decide to create their own test to use for an evaluation, but this task is more difficult than it appears. If this is the chosen strategy, it would be very helpful to partner with a measurement expert who can provide advice to the team.	Because it is locally developed, the test is aligned to both state and local standards. No national or state norms, so DLP student performance cannot be compared to populations.
Locally developed translations of mathematics assessments	The measurement community generally believes that test translation is inferior to assessments development in the language from "the ground up", but this might be a viable option for some DLP programs if resources are limited	Might be an infringement of copyright. Check with the test publisher. Translating an existing test is not a simple task. Can serve as a good professional development opportunity for teachers. No national or state norms.

Case studies of three school DLP evaluations. I have conducted dozens of evaluations of DLP programs, sometimes mathematics was part of the evaluation, sometimes not. The three described below all assessed mathematics knowledge, but for different reasons.

Case Study #1: Mann Elementary (all names are pseudonyms:

Wilson Elementary is a neighborhood attendance-based school located in a middle to upper middle-class neighborhood in a large urban area. About half the school's students, at the time of the evaluation, were native Spanish speakers (NSS). Most of the NSS lived in a large apartment complex near the school; their families were mostly working-class and many worked in semi-skilled jobs at a massive hospital and medical enterprise near the school. The other half of the students were native English speakers (NES) and lived in single-family homes surrounding the school. Many of the NES parents worked as skilled, medical professionals. Several of the NES parents were bilingual, and in a few cases, they had raised their children bilingually. The DLP program (Spanish/English) was optional and filled one class per grade level (of three to four total at the school).

This evaluation was driven partly by the need to report its success and challenges to a funding agency and thus required an "external" evaluation, which I conducted. The funding agency did not require an assessment of mathematics, but in order to convince the NES parents to participate in the DLP, they wanted to have their children's mathematics knowledge assessed in both English and Spanish (mathematics was taught in Spanish from grades K-2). The NES parents were mostly concerned that their children's achievement might not keep up with other students not in the DLP. The NSS parents had similar concerns but were not as vocal. I administered the *Logramos®* literacy and mathematics tests, and the parallel ITBS tests in English. The results demonstrated that the NES students were performing slightly below their non-DLP counterparts on Spanish mathematics in the early grades, but that by the time the DLP students were in 4th grade, they were performing at or above the non-DLP students in mathematics in English. In fact, the NES (middle class students) were outperforming NSS (working class students) in mathematics, even when the test was given *in Spanish*.

Case Study #2: Wheatley Elementary School

Wheatly is located in a town of approximately 50,000. The population is largely working class. The neighborhood surrounding the school is largely older single-family homes, many built in the 1920s, now rentals and many in need of repair. The compelling aspect of the DLP was that the NES students in the DLP were nearly all African-American, most of whom spoke a local version of African-American Vernacular English Dialect, The purpose of the evaluation was two-fold. First, it was grant funded and therefore required an evaluation. Second, the school leaders wanted to know if the DLP students would outperform students enrolled in the traditional early-exit bilingual program, also offered at the school. The evaluation found that the DLP NSS students learned Spanish literacy skills at rates that matched, and at times, exceed those in the traditional bilingual program. Also of interest was the Spanish achievement of NES students, in both mathematics and literacy. For this evaluation, I selected a random sample of students to take the *Batería III* and the *Woodcock-Johnson III*. A random sample of students was necessary because both these assessments require an individual administration, but also provide much more information than any of the group administered tests. DLP educators should know that effective evaluations can be done without assessing all the students, and in some cases, this is the preferred strategy. In any event, the evaluation found that the NES students needed more time (i.e., grade levels) to meet the mathematics achievement of their peers in an English only program. However, by 5th grade, the NES students were performing as well as their English only program peers, and, of course, they had strong Spanish language skills by this time.

Case Study #3: Rivera Elementary School

This evaluation at Riviera was motivated by a desire to learn which language learning program in a single district was most effective. I was asked to compare (a) an SEI program, (b) a “late-exit” program-students transitioned to all English instruction at about 3rd to 4th grade, and (c) a DLP program. Readers of this volume and those familiar with language program evaluation will recognize the assessment of these three programs as common to many studies in language education (e.g., The Ramirez Report [Ramirez, Yuen & Ramey (1991)]) The school district wanted to know which program was most effective in teaching English, but I convinced them to also assess the late-exit and DLP students in Spanish mathematics as well, hoping to find the cross-content transfer I mentioned earlier. I should point out that the DLP program did not reflect the desired mix of students’ native languages (e.g., half NES and half NSS) in the same classroom. In this case, the DLP students were over 90% NSS, thus they did benefit from NES models. Of course, the SEI students were not tested in Spanish mathematics, but the results indicated that the late-exit students and the DLP students performed equally in mathematics, in both languages, until the later grades when the DLP students moved well ahead of their late-exit counterparts in Spanish mathematics. Although it was not a strict research study, I believe these findings are evidence that the DLP students benefitted from cross-content transfer. These findings were sufficient to convince the district that the DLP was effective and merited continued support and resources.

General suggestions for evaluating mathematics in DLP programs

1. **If you do not intend to analyze it, do not collect it.** Far too many educators believe that assessing students indicates that they are "data driven" and care deeply about student achievement. The burden for this over-assessment falls on teachers (and the students), who are apt to grow cynical of all the testing when no one uses the results for any meaningful decisions. Do not assess students unless you (a) have a genuine purpose in mind and can justify that purpose to teachers and other stakeholders (including the students), (b) are willing to analyze the results, (c) share the findings with your stakeholders in an open forum, and (d) invite everyone to consider how the results can be used to make important decisions about the curriculum, instruction, or programming.
2. **Be prepared to find class differences, regardless of native language.** I have found in my evaluation of DLP programs that native language is a poor predictor of mathematics achievement. Students from middle and upper-class backgrounds—whose parents typically have more education—begin to outperform working class students in *both* languages beginning at about 3rd or 4th grade, in both literacy and mathematics, but mathematics in particular. As I noted in Sample Evaluation #1, many of the NES students began to outperform their

NSS peers in Spanish version tests of mathematics as early as 3rd grade. By 5th grade, the difference was even more pronounced. At first, the DLP leaders thought I had incorrectly analyzed the data. They asked, "How could NES students do better than NSS on a test of mathematics *in Spanish*?" After making sure I had not made any mistakes, I explained that part of the reason is that they had done their jobs well. The NES students learned Spanish with great success. Accounting for all the reasons they outperformed the NSS would suggest I have a theory to explain why we find myriad class differences in educational achievement, especially on tests of achievement. I do not, but do not be surprised if you find them in your DLP evaluation.

3. **Present your findings at professional meetings and engage teachers to the extent possible.** As Naqvi, Schmidt and Krickhan (2014) point out, the best route to teachers who understand linguistic transfer and other DLP features is the development of professional learning networks where teachers can participate and collaborate in research and evaluation.
4. **Consider using more participatory evaluation methods to evaluate DLP**. Because it is grounded in the experience of administrators, teachers, parents, and students, participatory evaluation is more likely to provide information that is useful to program administrators and policy makers (Brisolara, 1998). Participatory evaluation often results in positive changes within an organization or project, including increased communication between staff members, beneficial effects on program development, and higher quality evaluations (Upshur, &

Barreto-Cortez, 1995). But do not hesitate to enlist the help of professionals.[3] Program evaluation is a specific discipline that requires training.

5. **Use caution in interpreting achievement scores for students under the age of 9.** Young children are notorious for providing unreliable data. It is fine to assess students at these ages (most tests offer versions for students as young as Kindergarten age), but use the results with caution.
6. **Use the right units when comparing groups.** Often, the benchmarks used in school accountability are not useful in a program evaluation. For example, in a published evaluation which admittedly addressed many more variables than mathematics student achievement alone, the researchers found that the students at an established DLP had a higher *percentage* of students passing a standardized examination in mathematics (tested in English) than the state average or other students in the same district. Good evidence, but in order to make a valid claim, they should have used *scaled scores* rather than percent passing. Those mean (average) scores should have then been compared to other students in the district and matched by class (e.g., free and reduced-price lunch).
7. **Mathematics instruction in the LOTE will end, and DLP educators must plan for the transition.** There is no school system in the US that allows students to take all their mathematics courses required for graduation in a LOTE. We may claim that this is a flawed plan, but it is unlikely to change. Therefore, mathematics instruction in the LOTE, which is typically a wise idea (recall our

[3] A strong evaluation can come only from the result from collaboration of an entire DLP team, which might include the use of an expert program evaluator.

discussion of cross-linguistic and cross-content transfer), must be assessed carefully so that DLP educators can plan for the inevitable move to instruction in English.

In conclusion, I hope that DLP educator will consider mathematics, both its instruction (not just the language of instruction) and assessment, more carefully. I further hope that this chapter has promoted a new understanding of the importance of assessing mathematical knowledge in two languages, as well as provided some tools and advice. Conducting a rigorous and valid program evaluation of a DLP program is no “mean” feat, but done well it is a worthwhile endeavor that can inform the work of administrators, teachers, parents, and other school personnel. DLP are a leap of faith, but with a strong evaluation component, the leap can be made with confidence.

References

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*(3), 219-234.

Brisolara, S. (1998). The history of participatory evaluation and current debates in the field. In E. Whitmore (Ed.), *Understanding and practicing participatory evaluation* (pp. 25-41). San Francisco: Jossey-Bass.

Celedón-Pattichis, S., Musanti, S. I., & Marshall, M. E. (2010). Bilingual elementary teachers' reflections on using students' native language and culture to teach mathematics. In *Mathematics Teaching and Learning in K-12* (pp. 7-24). Palgrave Macmillan US.

Chen, F., & Chalhoub-Deville, Micheline. (2016). Differential and long-term language impact on math. *Language testing, 33*(4), 577-605.

Cheng, L., Li, M., Kirby, J. R., Qiang, H., & Wade-Woolley, L. (2010). English language immersion and students' academic achievement in English, Chinese and mathematics. *Evaluation & Research in Education, 23*(3), 151-169.

Common Core State Standards (2014). Retrieved from https://www.cde.ca.gov/be/st/ss/documents/ccssmathstandardaug2013.pdf, October 31, 2019

Creighton, T. B. (2006). *Schools and data: The educator's guide for using data to improve decision making*: Thousands Oaks, CA: Corwin Press.

Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review, 34*(4), 344-377.

Cummins, J. (2008). Teaching for transfer: Challenging the two solitudes assumption in bilingual education. In *Encyclopedia of language and education* (pp. 1528-1538). New York: Springer.

Danesi, M. (2003). Conceptual metaphor theory and the teaching of mathematics: Findings of a pilot project. *Semiotica*, 145, 71–83.

De Jong, E. J. (2002). Effective bilingual education: From theory to academic achievement in a two-way bilingual program. *Bilingual Research Journal, 26*(1), 65-84.

del Valle, J. (2000). Monoglossic policies for a heteroglossic culture: Misinterpreted multilingualism in modern Galicia. *Language and Communication, 20*(2), 105-132.

Dewey, J. (1938.) *Logic: The theory of inquiry*. Published as Volume 12 of *The Later Works (1925-1953)*, J.A. Boydston (Ed.). Carbondale, IL: Southern Illinois Press

Fernandez, C., & Yoshida, M. (2012). *Lesson study: A Japanese approach to improving mathematics teaching and learning*. New York: Routledge.

Halliday, M. A. (1993). Towards a language-based theory of learning. *Linguistics and Education*, *5*(2), 93-116.

Lindholm-Leary, K., & Hargett, (2006). Evaluator's toolkit for dual language programs. Washington, DC: Center for Applied Linguistics. Retrieved 3/23/2019 http://www.cal.org/twi/EvalToolkit/downloads.htm#evaluator

MacGregor, M., & Price, E. (1999). An exploration of aspects of language proficiency and algebra learning. *Journal for Research in Mathematics Education*, 30, 449–467.

Morrison, J. Q., & Harms, A. L. (2018). *Advancing evidence-based practice through program evaluation: A practical guide for school-based professionals*: Oxford University Press.

Moschkovich, J. (2012). Mathematics, the Common Core, and language: Recommendations for mathematics instruction for ELs aligned with the Common Core. *Commissioned papers on language and literacy issues in the Common Core State Standards and Next Generation Science Standards, 94*, 17.

Moses, R., & Cobb, C. E. (2002). *Radical equations: Civil rights from Mississippi to the Algebra Project*: Beacon Press.

Naqvi R., Schmidt E., Krickhan M. (2014). Evolving 50%-50% bilingual pedagogy in Alberta: What does the research say? *Frontiers in Psychology*, 5, 413.

National Assessment of Educational Progress. (2009). *2013 Mathematics and Reading Assessment Reports*. Retrieved January 15, 2016 from http://www.nationsreportcard.gov/reading_math_g12_2013/#/

National Research Council. (2001). Adding it up: Helping children learn mathematics. J. Kilpatrick, J. Swafford, & B. Findell (Eds). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Perry, R.R. & Lewis, C.C. (2011). *Improving the mathematical content base of lesson study summary of results*. Retrieved January 24, 2016 from http://www.lessonresearch.net/IESAbstract10.pdf.

Pimm, D. (1987). *Speaking mathematically: Communication in mathematics classrooms*. New York: Routledge.

Ramirez, J. D., Yuen, S. D., & Ramey, D. R. (1991). *Longitudinal study of structured English immersion strategy, early-exit, and late-exit transitional bilingual education programs for language-minority children*. San Mateo, CA: Aguirre International.

Rolstad, K., Mahoney, K., & Glass, G. V. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy*, 19(4), 572.

Snow, C. E. (2010). Academic language and the challenge of reading for learning. *Science*, 328(5977), 450–452.

Spanos, G., Rhodes, N. C., Dale, T. C., & Crandall, J. (1988). Linguistic features of mathematical problem solving: Insights and applications. In R. Cocking & J. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 221–240). Hillsdale, NJ: Lawrence Erlbaum.

Sugarman, J., Arteagoitia, I., Coburn, C., Gallagher, C., Montee, M., & Schissel, J. (2007). Spanish-language assessments for dual language programs. *Center for Applied Linguistics*. retrieved http://www.cal.org/twi/pdfs/assessments.pdf

Tellez, K. (1997). Class placement of elementary school emerging bilingual students. *Bilingual Research Journal*, 22, 279-295.

Téllez, K. (2010). Teaching English language learners: Fostering language and the democratic experience. In: Boulder, CO: Paradigm Publishers.

Téllez, K., & Mosqueda, E. (2015). Developing teachers' knowledge and skills at the intersection of English language learners and language assessment. *Review of Research in Education, 39*(1), 87-121. doi:10.3102/0091732X14554552

Téllez, K., Flinspach, S., Waxman, H. (2005). Resistance to scientific evidence: Program evaluation and its lack of influence on policies related to language education programs (pp. 57-76). In R. Hoosain & F. Salili (Eds.), *Language in multicultural education*. Charlotte, NC: Information Age Publishing.

Téllez, K., Moschkovich, J. N., & Civil, M. (2011). *Latinos/as and mathematics education: Research on learning and teaching in classrooms and communities*: Charlotte, NC: Information Age Pub Incorporated.

Upshur, C. C., & Barreto-Cortez, E. (1995). What is participatory evaluation (PE) and what are its roots. *The Evaluation Exchange: Emerging Strategies in Evaluating Child and Family Services, 1*(3/4). Retrieved March 10, 1998 from the World Wide Web: http://gseweb.harvard.edu/~hfrp

<u>Appendix</u>

Below are two activities for using the chapter's tool: *List of assessments suitable for evaluating mathematics learning in DLP, including descriptions, benefits and disadvantages of each* (Table 1).

Activity 1:

As you explore the assessments listed in the chapter's tool (see pages 14-15), think about which assessments would be appropriate for the two, different DLP contexts shared below. Review each of the descriptions and consider which assessments would be the most appropriate for each context. Use the guiding questions below each description to refine your recommendations.

<u>Escuela Mundial</u>
Escuela Mundial teaches mathematics in Spanish until the 4^{th} grade. In their final year of the program, the students, as fifth graders, transition to mathematics in English. This transition was thought necessary because the middle school where the students all attend did not have a teacher who could teach mathematics in Spanish. However, they recently hired a mathematics teacher who attended secondary school in Mexico City and feels confident in her ability to teach mathematics in Spanish (she holds the state's bilingual authorization/credential). Although the middle school could possibly now offer mathematics in Spanish at the sixth grade, seventh grade mathematics will be in English. The teachers, administrators, and parents want to know the balance of students' knowledge of mathematics knowledge in both languages, so they can decide if continuing in Spanish mathematics at the sixth grade is a sound idea. The school district is willing to pay for the cost of the assessment.

Guiding Questions:
1. Which test or tests from the tool might be used to make a decision about when to transition from mathematics instruction in Spanish to instruction in English?
2. Which students should be assessed? Just the Escuela Mundial students?
3. Which grade levels should be assessed?
4. How might the school share the data with the school community?
5. What evidence would convince the parents that the school district is making the correct decision?

<u>Escuela Dos Alas</u>
Escuela Dos Alas just adopted a new Spanish mathematics curriculum. Mathematics is taught in Spanish in grades K-3, transitioning to English in grade 4. The teachers and administrators want

to know if the new curriculum is teaching the CCSS (Escuela Dos Alas in located in California) sufficiently. Students in the 3rd to 5th grade have taken the Smarter Balanced Assessment (in English only) for the last 5 years, so the school has some longitudinal data. Because the school and district have no additional funding to purchase additional tests, they must rely on the existing assessments.[4] The school also wants to know how much, if any, mathematical content knowledge is lost when students make the transition from Spanish mathematics to English.

Guiding Questions:
1. Which test or tests from the tool might be used to determine if the new curriculum is teaching the CCSS sufficiently? (Recall that no additional funding is available.)
2. Should the school take on the task of development assessments in Spanish? If so, how can they show that their assessments are valid?
3. Does the curriculum come with summative assessments in Spanish? If so, how might they be used?

Activity 2:

Finding the best assessment for your program: Using the Mental Measurements Yearbook

While not well known by many educators, assessment professionals are very familiar with a decades-long effort to catalog and review educational and psychological assessments. Currently known as the Buros Center for Testing, it has been housed at the University of Nebraska since 1938, and has reviewed over 10,000 tests since its inception, which have been published in volumes known as Mental Measurements Yearbook (MMY).

Each assessment is reviewed by two independent reviewers, who are not on the payroll of the publisher or have used the assessment in their research. It is important to note that not all tests are reviewed, but many of the most commonly used educational assessments are included. Of specific interest for DLP educators, the center has published two volumes (thus far) of *Pruebas Publicadas en Español*, a parallel volume to the MMY.

To begin searching, go the search engine for the MMY (https://marketplace.unl.edu/buros/) You'll find several search terms. For the purposes of explanation, type in "Aprenda" and you'll find that a review of the assessment, *Aprenda®: La prueba de logros en español—Segunda edición. A*s the site advises, the $15.00 fee buys you the review of the test, not the test itself. Alternatively, most university libraries have a reference copy of the MMY. If you can find a copy of the MMY in print, you will not need to purchase the review online. A quick visit to the online catalog will let you know if your local university library has a copy of the MMY. If you are close to a university library, I would suggest you make a visit and talk to the reference librarian. They are typically very helpful in finding the MMY and helping to find a review of a specific test. With respect to the specific review of the *Aprenda®,* you will find it would be a very useful assessment of mathematics in Spanish and thus a good tool for DLP educators.

Next, try searching using various search of interest to you. Try to find other assessments that would be useful in a DLP context. If you are a teacher working in a school district, you might wish to search for the assessments required by district leadership. Do not be surprised to find that the assessments used in your school or school district are not considered worthy by the experts. I cannot tell you what to do when you find a poor assessment being used in your school, but I do believe that educators have a professional obligation to use valid and reliable assessments to make important instructional decisions. I can also say that most district curriculum leaders do not have a background in tests and measurements and have never heard of the Buros Center or the Mental Measurements Yearbook reviews.

I admit that searching test reviews in the MMY is a bit complicted—and expensive if you have to buy the reviews—but I can say that schools and school districts can spend many thousands of dollars purchasing tests that do not answer the questions they seek. I hope by having the MMY in your DLP evaluation toolkit, you will be the smart teacher or administrator who helps to make a good assessment choice or at least avoid a poor decision.