

Similarities and Dissimilarities in the Teaching of English and French
in a Dual Language Instructional Setting

Robert Savage * ¹ Adriana Pace

*Corresponding author

University College

London

WCIH OAA

* ORCID: orcid.org/0000-0002-5645-6002

Word count = 7300

¹ Adriana Pace is now at Sir Wilfrid Laurier School Board, 235 Montée Lesage, Rosemère, QC Canada, J7A 4Y6.

Abstract

This study investigates the relationships between the observed quality of classroom teaching assessing Atmosphere, Instruction, Management and Student Engagement, (AIMS, Roehrig & Christesen, 2010), and growth in reading and listening comprehension across grade 4 in typical English-French dual language learners. For the majority, (67%), English was a first language (L1), for 16% French was an L1. $N = 332$ children in 36 classrooms were assessed on reading and listening comprehension measures and observed in the fall and winter semesters in English and French. Results showed that, overall, ratings of AIMS sub-scales did not differ markedly across English and French classrooms. Classroom teaching quality ratings were not, however, correlated across English and French classes. AIMS ratings in either language for Atmosphere were reliable and relatively stable between fall and winter, as was Management in English. AIMS ratings predicted growth in classroom-level variance in reading and listening for English and French both concurrently and longitudinally. These results suggest that there are broad, stable commonalities in effective teaching in English and French classrooms that can be validly and reliably observed. Practically, such observations of (majority) second language teaching quality can facilitate assessment of student needs and mentoring and development needs of teachers.

Similarities and Dissimilarities in the Teaching of English and French in a Dual Language Instructional Setting

A body of recent educational research focuses on identifying observable characteristics of ‘effective teaching’. These studies are often concerned with relatively broad features of practice such as the atmosphere and tone, the curriculum delivery, differentiation, and classroom management style of classrooms (see e.g. Deault & Savage 2013; Pressley, 1998). There has however been relatively little research on whether these broad characteristics of effective classrooms generalize to the same children when receiving what is for the majority of them, their L2 instruction (and where the same students, by definition, are often operating at lower levels of proficiency and experience).

Whether these broad features of effective teaching approaches are similar for the same children in dual language instructional contexts is relevant to a number of theoretical and practical concerns. These include: The implicit universality of such claims; the sometimes disputed claims for dual language teaching models to be evidence-based or appropriately differentiated; the ability to use validated classroom observation to audit, mentor, and facilitate excellent teaching in dual language contexts; the subsequent refinement of observational tools for dual language contexts. It is this gap in research on broad observable similarities in the teaching of L1 and L2 and their impacts on outcomes in a dual language instructional context that is undertaken here. Here, the broad features of effective literacy teaching in L2 contexts rather than attainment more generally-construed is investigated. This issue is important as the literacy needs of children in L2 classrooms are of particular concern (e.g. August & Shanahan, 2006; Savage & Côte, 2017). To this end we first explore research evidence on effective teaching more broadly construed and then

more specifically on literacy, both drawn primarily from monolingual instructional contexts, before considering research in effective dual language literacy teaching.

Research on Effective Teaching

There exists a significant body of work that attempts to establish the broad characteristics of effective teaching by relating measured growth in attainment to classroom observations of teaching practice. Using the CLASS observational tool, Pianta, Ia Paro, and Hamre, (2007), have suggested that 3 broad categories of teaching practice are important: Emotional Support, Classroom Organisation, and Instructional Support. While all three features are important, they identify higher levels of emotional support, as measured through the warmth of adult-child interactions and responsiveness to children's needs as key factors that affect growth in achievement (Hamre & Pianta, 2005; Pianta et al., 2007). Comparable findings concerning the range of general pedagogical, curricular, managerial, and inter-personal features of effective teaching have been reported by others using a range of observation tools (e.g. using AIMS, Pressley, 1998, Pressley et al., 2001; using the CLOS, Louden et al., 2005). Such observations are important for a future science of effective teaching, and practically, they can also validly serve an accountability and professional development function. These observational studies however often reflect a snapshot in time, whereas teaching may vary across time. There is thus a need for the evaluation of such observational measures undertaken more than once over the course of a school year. Additionally, tools such as CLASS and CLOS were not specifically developed for assessing literacy teaching but teaching more generally, so may not pick up on attributes of effective and balanced teaching suggested by other lines of research evidence (e.g. National Reading Panel, 2000; Pressley et al., 2001; Savage & Cloutier, 2017).

Research on Effective Literacy Teaching

A distinct approach taken by researchers on effective literacy teaching has been to look at the characteristics of literacy teachers nominated as being excellent practitioners (Pressley et al., 2001, Topping & Ferguson, 2005). Observation of the classroom literacy instruction practices of such teachers suggest it is marked by the use of explicit strategy instruction, and the regular immersion of children in reading and writing experiences based on exposure to quality children's literature. Excellence in classroom management was clearly evident, and was characterised by the promotion of student self-regulation over authoritarian or teacher-directed approaches. A clear sense of community and democracy in classroom decision-making was also evident. Expert scaffolding and the matching of curriculum content to student competencies also occurred frequently. Finally these effective teachers promoted links outside the classroom, supporting diversity in strong family – community links, (Pressley, 1998; Pressley et al., 2001). Pressley et al. (2001) suggest that no one component of instruction characterizes effective teaching, rather that effective grade one literacy classrooms included a multitude of effective practices.

Built upon this and related work, the Classroom AIMS Observation Instrument was developed to formalize observation of the quality of classroom Atmosphere, Instruction, Management and Student Engagement, (AIMS; Roehrig, & Christesen, 2010). Roehrig and Christesen present evidence that the AIMS tool is a psychometrically valid and reliable observational tool when used to determine effective literacy practices among K-12 classroom teachers. They also claim that AIMS can be applied to determine areas for professional development, and to facilitate self-reflection among both current and pre-service teachers. Roehrig and colleagues present evidence that improvements in the effectiveness of instructional

practices in beginning teachers can be nurtured through mentoring techniques that incorporate AIMS data to assess professional change (Roehrig, Bohn, Turner, & Pressley, 2008).

Finally, there is evidence that there exists no ‘one size that fits all’ for the constituents of maximally effective teaching. Deault and Savage (2013) used a nested hierarchical design to assess different aspects of literacy teaching as predictors of change in students’ reading and attention in first grade in English (the L1). Observations of literacy teaching were obtained for 18 classrooms using the *Classroom AIMS Instrument* (Roehrig & Christesen, 2010). Analyses showed significant ability by classroom interaction effects: For students who started grade 1 with relatively strong reading skills, classroom management predicted higher rates of growth in reading comprehension. That is, for stronger readers, greater observed independence and self-regulation in classrooms was associated with greater growth in attainment. On the other hand, for students with weaker initial reading ability at the start of Grade 1, student engagement, in tasks supported by teachers predicted greater reading comprehension growth. Deault and Savage thus concluded that these results are consistent with ‘goodness-of-fit’ models of the effect of classroom practices on children’s reading wherein the optimality of teacher practices in literacy instruction depend on the current attainment of particular students (see also Connor et al., 2009, 2011).

Research on Dual Language Teaching

Research on additional and dual language instruction has for some time been diverse and multidisciplinary in nature (see August & Shanahan, 2006; Genesee & Jared, 2008; Mansour & Butler-Kisber, 2015 for reviews). August and Shanahan for example overviewed research on basic processes in L2 acquisition including the central role of L1 language skills, while also drawing attention to the variation in performance, the poor quality of assessments (then)

available, issues concerning the interpretation of low performance, the role of home language and the relatively modest amount of work on pivotal social and cultural variables in L2 acquisition. A review of theoretical models of L2 teaching is beyond the scope of this paper but Cummins (2014) identifies the need for effective L2 teaching to scaffold meaning, connect to student lives, active prior knowledge, affirm identity, and extend language use beyond the classroom.

Research on Dual Language Literacy Teaching

There is much recent research on the extent to which literacy-relevant linguistic sub-processes that are impacted by teaching such as phonological awareness, decoding and listening comprehension, overlap or predict first language (L1) and second language (L2) in dual language contexts (August & Shanahan, 2006; Genesee, Geva, Dressler, & Kamil, 2006; Melby-Lervag & Lervag, 2012; Savage & Pace, 2017). In terms of research on L1 oral language skills, analyses of processes such as the use of cognates shared across languages, or the use of ‘bootstrapping’ skills of L1 competencies on L2 comprehension suggests they are important aids to effective language teaching (August & Shanahan, 2006; Hamayan, Genesee, & Cloud, 2000).

There is also research on candidate oral language micro-processes in L2 literacy teaching. For example, Lyster, Collins, and Ballinger (2009) used a bilingual read-aloud project where both English and French teachers read from a common storybook in each language, to explore whether such practices raise teachers’ awareness of the bilingual resources of their students, and whether it engender both teacher and student cross-linguistic collaboration in three classes of six- to eight-year-old children. Results of qualitative focus-group analyses suggested enthusiasm of both students and teachers for the project, the opportunities it created for teachers and students to focus on both language and content, and their collaborative approach to this enterprise. Lyster, Quiroga,

and Ballinger (2013) explored the effects of a derivational morphology-focused intervention with shared books on outcome measures of morphological awareness administered both in English and French. Results suggested that morphological ability improved in both languages following intervention (see also Gunning, White, & Busque, 2016). It is also quite clear from this review that there is currently a rich and complex research base on language processing in immersion-specific pedagogy with its own particular and unique knowledge base and approaches.

The rationale for the present research

The existing research suggests many specific practices that might be used to teach dual literacy effectively, a context where by definition many of the same students have varying levels of exposure to L1 and L2. There has however been relatively little attention to the sorts of broader processes might be involved in studies of dual language literacy teaching comparable to those undertaken in monolingual instructional contexts described earlier. Such analyses of broad features of teaching such as Atmosphere, Management, and Instruction in dual language contexts are however relevant to a number of important theoretical and practical questions. We thus describe these below.

Assessment of ‘transmission’ models of education. A major issue in dual language research concerns the extent to which teaching practices are optimal or evidence-based in nature (Cummins, 1998, 2007a, 2007b, 2014; Savage & Côte, 2017). Most notably, Cummins has disputed the common process of teaching L1 and L2 separately, rather than valuing and synergistically using L1 capacities. He terms this former approach a ‘two solitudes’ model. Alongside this model Cummins (1998) noted that L2 immersion teaching has been characterized as using a highly teacher-centered or ‘transmission-oriented’ approach where teachers deliver the

curriculum with minimal opportunities to use L2 in creative or problem-solving tasks. Cummins (2007a) describes some examples of negative impacts in L1 and L2 contexts of changes in U.S. educational policy that compelled teachers to use transmission-oriented approaches.

We can find no empirical data from studies involving direct observation of classrooms to confirm or refute this long-standing (and informally, oft-repeated) claim in French L2 classes in the Canadian context. We seek therefore to evaluate its validity in a reasonable number of representative regular classrooms. Quite specific hypotheses can be constructed using the AIMS tool. If the teaching of French in French immersion classes is indeed dominated by teacher-led ‘transmission’ modes of instruction, then in such contexts, while I (Instruction) *may* be comparable in classrooms A (Atmosphere) and M (Management) with their focus on democratic engagement, community, and student self-regulation respectively will be lower than in English language classrooms for the same children.

2. Assessment of ‘goodness-of-fit’ models. As noted earlier, recent research on monolingual instructional contexts literacy growth has suggested that for differing levels of ability, different aspects of quite broad aspects of effective teaching such as Atmosphere, Instruction, Management and Student Engagement might be relevant at this grain of analysis. For example, greater self-regulation of learning was suggested for the initially more capable, and more teacher-directed sustained attention appeared to mark growth for the initially less capable (Deault & Savage 2013). Given that L2 experience and (often) performance is by definition lower than L1 in elementary school students this might predict different AIMS predictors for the same children when followed in classrooms, which, for the majority of students are L2 classrooms. This hypothesis might predict greater degrees of teacher-led aspects of Instruction, but lower levels of Atmosphere and Management in French L2 over English L1 classes.

3. Assessment of implicit claims to universality. Existing effective teaching practices might be seen as ‘universals’ across both languages irrespective of children’s absolute levels of attainment. Certainly they have not been *explicitly* limited to monolingual teaching contexts by their authors. We know of no published evidence on this particular question to date, so we seek to test these implicit claims to universality. Against that claim, it may be that different broad teaching approaches are needed given the motivational and socio-cultural differences in L2 exposure relative to an L1 (e.g. August & Shanahan, 2006), and the fact that French L2 teaching involves a high degree of immersion-specific pedagogy. The testing of a universality claim does not however in any way devalue (or even attempt to directly measure) this complex L2 pedagogy. Rather it evaluates the view that there is a common ‘general instructional expertise’ (Cummins, 2007b) shared in L1 and L2. This version of a universality claim predicts that the same features (all 4 elements of AIMS) will thus predict improvement in both English and French longitudinally.

4. The use of observational tools to audit and facilitate teaching practice. Practically another reason why it is important to compare the broad characteristics of observational tools in dual language contexts relates to comprehensive assessment. Consider popular Response-to - Intervention (RtI) models for supporting literacy and assessing literacy difficulties (e.g. Jimerson, Burns, & van der Heyden, 2016). In RtI, children’s literacy difficulties are evaluated using a test-teach-re-test model. Here support for children commences with evidence-informed regular classroom teaching. Increasing levels of ‘intensity’ as required i.e. from whole class to small, group or one to-one tutoring are added as need persists. If literacy difficulties remain a child may be considered eligible for special educational support on the assumption that their literacy difficulties are constitutional in origin. This eligibility for special educational support assumes of

course that the regular teaching experienced was of high quality, and indeed, eligibility may be dependent on formal accounting of the quality of pedagogy. Another relevant domain is allied high stakes decisions made in some dual language contexts to remove a low-achieving child from immersion contexts and place them in an L1 only curriculum. In both cases evaluation of possible ‘within-child deficits’ are complimented by observations of the effectiveness, suitability, and quality of regular L2 teaching they receive (Genesee, Savage, Erdos, & Haigh, 2013). In addition, formal validation of tools like AIMS in L2 contexts opens up the possibility of using them for mentoring and professional development work in a manner akin to that undertaken in L1 (Roehrig et al., 2008). Furthermore, given the growing evidence on the potential of cross-linguistic collaboration (Lyster et al., 2009; 2013; Gunning et al., 2016) we sought to explore the existent degree of similarity in teaching of pairs of teachers within the same school using AIMS. Alongside the earlier articulated aim concerning establishing the consistency of AIMS factors over time, the present study explores these issues above here for the first time.

Materials and Method

Participants

Teachers. Participating pairs of English and French grade 4 teachers were recruited by Language Arts consultants in their respective schools boards in Quebec, Canada. Schools came from 3 school boards representing one urban and two suburban districts. The final sample ($n=36$ classrooms) included 33 women and 3 men (thus 18 pairs of English language arts with French as a second language teachers) with varying levels of teaching experience ranging from first-year teachers to almost-retired teachers. Participating teachers then chose one of their shared grade 4 classrooms from which data would be collected.

Students. Recruitment was undertaken over two school years. In the first year, consent letters were initially sent to all of the parents of the children in the classroom without exclusion ($n=246$). Of these, consents for 206 children were received. Similarly, in the second year of the study, consent letters were again sent out to all parents of children, without exclusion ($n=143$). Of these, consents for 126 children were received. A consort flow diagram depicting all phases of participant recruitment and involvement is depicted in Figure 1. The final attained sample ($n=332$) included 166 boys and 166 girls with a mean age of 9 years 0 months. Children who participated in this study were all in grade 4 classrooms and enrolled in bilingual or immersion programs where they received both English and French instruction throughout the school year. Dual language instruction was structured either on a daily basis (where children received both English and French half-day instruction) or on a weekly basis (where children received instruction one day in English and the next in French) and had been in place throughout the children's school careers to date.

Under provincial language laws generally only the children or siblings of children, educated in English, whose parent or parents received their own education in English may typically attend such dual language immersion schools. Thus the home language of children in these schools is likely to include substantial amounts of English. However there are a number of exceptions to this rule for temporary workers, students with learning disabilities and families in acute crises, amongst other categories. Therefore a parent questionnaire was sent home along with the consent letters in an effort to gather more direct information about children's home language backgrounds. Questions explored mother and father's native languages and languages spoken at home between child and each parent (English, French and other if applicable). Overall, 224 participants (67%) identified English as their first language, 56 (16%) identified French as their

first language, 26 (8%) listed “other” (subsequently referred to as ‘Allophone’), including a range of languages such as Urdu, Mandarin, and Arabic. The remaining participants 26 participants (8%) did not complete the parent questionnaires.

Measures

The *Group Reading Assessment and Diagnostic Evaluation (GRADE)*, Form A and B is a standardized group-administered, reading test which also assesses listening comprehension (Williams, 2001). Reading for meaning is the fundamental purpose of reading so reading comprehension is the broadest and most useful index of performance. We also assess listening comprehension as it is both a component of reading comprehension, and a relatively pure measure of linguistic processing underlying reading comprehension (Kirby & Savage, 2008). It was not possible to administer a word reading or fluency measure as educational administrators involved in the project opposed their use in this work.

Students were assessed on two of the four subsections contained in the GRADE, listening comprehension and passage comprehension. The GRADE was selected for this study because it was designed to provide a precise measure of change in comprehension skills, because it assesses reading and listening comprehension. Testing typically takes between 60-90 minutes and is completed in a whole-class format. Reviews of the GRADE (Fugate, 2003; Waterman, 2003) indicate that it is a reliable and valid measure of children’s reading ability.

Due to the dual-language nature of this study, researchers sought to assess both English and French reading skills in students and thus, one of the versions of the GRADE (Form A) was translated. The French version was directly translated from the English version but was simplified somewhat by the Language Arts consultants. Here, grammar and verb tense was simplified

somewhat to better assess grade 4 students receiving French as a second language instruction in French immersion or bilingual schools. The tests were designed to assess children on various skills including their vocabulary, grammar, idiom, inferential, and non-literal understanding of items. Questions included present and past verb tenses. To the greatest extent possible, the consultants tried to maintain a similar verb tense in both the listening and reading comprehension items from English to French. However due to its complexity, one of the most advanced passages included in the reading comprehension section of the GRADE was omitted and replaced by a French text provided by one of the consultants. Copyright permission was sought and granted by the test publishers to allow translation of a limited number of copies of the standardized test.

Listening Comprehension. The *Listening Comprehension* subtest of the GRADE test contained 17 items. Students listened while the research assistant read aloud a sentence or group of sentences corresponding to one of four pictures in the student booklet selected by the child. This section took approximately a half hour for students to complete. The Spearman-Brown split-half internal reliability coefficient for the English measure in our sample in the fall semester was .46. The Spearman-Brown split-half internal reliability coefficient for the French measure in our sample in the fall semester was .48.

Passage Comprehension. The *Passage Comprehension* subtest contained 28 items. Here the students were required to read six different passages and answer four to five multiple-choice questions for each corresponding passage independently. This subtest took approximately one hour for students to complete. The Spearman-Brown split-half internal reliability coefficient for this English measure in our sample in the fall semester was .83. The Spearman-Brown split-half internal reliability coefficient for this French measure in our sample in the fall semester was .78.

Classroom observations. The classroom observations were measured by scoring elements of the Atmosphere, Instruction, Management, and Student Engagement subscales found in the classroom, using the (AIMS) questionnaire (Roehrig & Christesen, 2010). The content of this measure was drawn from the characteristics of teachers nominated as effective across studies described earlier (e.g. Pressley et al., 2001), inductively categorized to create four categories (Atmosphere, Instruction, Management, and Student Engagement).

As defined by the AIMS, Atmosphere refers to the physical and interpersonal environment of the classroom, encompassing attributes such as a sense of community, interest, focus on student effort rather than performance, opportunity for student choice, emphasis on the value of learning, high expectations for all students and use of informative feedback. Instruction/Content refers to the lessons and activities included in the literacy program, as well as the teacher's instructional style. Subcategories included in this domain reflect elements such as the degree to which the literacy content and activities are engaging for students, the density of instruction, cross-curricular connections, modeling and explicit teaching of thinking processes, scaffolding setting an appropriate level of challenge for individual students, provision of academic monitoring and encouragement of academic self-regulation. Management refers generally to the organization, rules, routines and procedures that guide the running of the classroom environment, including teachers' use of monitoring for on-task behavior and promotion of behavioral self-regulation. Student Engagement is defined as observable indicators of student engagement, including participation, excitement, and staying on task (Deault & Savage, 2013). See appendix A for further exemplification. The published reliability of the AIMS instrument is strong for each category: Atmosphere ($\alpha=.87$); Instruction ($\alpha=.90$); Management ($\alpha=.74$); and Student Engagement ($\alpha=.79$); (Roehrig & Christesen, 2010). In terms of test validity, and as reviewed

earlier, Deault and Savage (2013) report that AIMS was a significant predictor of reading and listening comprehension from fall to winter in Grade 1 students.

Procedure

The research design involved two distinct phases of data collection: Time 1 (t1) and Time 2 (t2) comprehension assessments each paired with fall and spring classroom observations. Graduate research assistants with backgrounds in educational psychology administered the GRADE and the AIMS classroom observations.

The *Group Reading Assessment and Diagnostic Evaluation* (GRADE) measure was administered to the whole-class. The English version of the GRADE, Level 4 Form B and the modified French version, Level 4 Form A were administered on two separate days to avoid student fatigue. Scores from the GRADE measure were obtained twice during the academic year (fall and winter) in order to assess growth in ability and to explore models of change in comprehension. The range of time between fall and spring assessments was typically around 6 months. All consenting children participated in whole-class assessment sessions where the research assistant and classroom teacher were both present.

Classroom observations using AIMS took place twice during the school year in the classrooms of the participating English and French teachers for approximately one and a half hours. This duration of observations is deemed sufficient to obtain reliable and valid data (Roehrig & Christesen, 2010). Teachers were informed that the observations were confidential, and that they were free to refuse consent for classroom observations. None chose to do so. Teachers were asked not to alter their regular teaching practices in any way. The range of time between fall and spring semester observations was typically 4.5 months.

Training of all observers was undertaken through exposure to- and review of-, the *Classroom AIMS Instrument* and the background research on effective teaching. Research assistants also practiced completing classroom observations by viewing videotapes of elementary classrooms and then further discussed and clarified item meanings on individual AIMS ratings until an inter-reliability of at least .7 was achieved across observer pairs. In the field, pairs of trained observers took detailed parallel field notes of their observations while independently observing a given language lesson. Following the classroom observation, each of the members of the pair of observers independently each scored the teacher and classroom according to the AIMS questionnaire. Observers rated each of the items on the AIMS subscales from 3-1 (3=Exemplary use; 2=Typical/inconsistent use; 1= Poor or rare use) depending on the degree to which they thought that item characterized the teacher and classroom.

Once each observer had completed the questionnaire by referring to their detailed field notes taken in the classroom, they then discussed each item individually in order to reach an agreed score. For example, if both observers had identically scored an item, then that score would count as the agreed score. For the items on which their scores were different (less than 7% of items overall), they then discussed their own observation and provided examples from their notes to support their score. They would then come to a consensus as to which score best “fit” the description of the observation of that item. The agreed scores were then entered in the SPSS database prior to analyses. Thus, the final product determined from the classroom observations was a set of four scores for each teacher, representing their average “agreed” ratings on the categories of Atmosphere, Instruction, Management and Student Engagement.

Results

Preliminary Data Analyses

Preliminary inspection of the data using conventional approaches (Tabachnick & Fidell, 2007) suggested that there were no marked kurtosis or skewness in the student-level pre and post-test GRADE scores in English and French. There were no clear outliers and, therefore, no data was excluded from analysis at this point.

The pattern of student-level missing data was inspected using a Missing Values Analysis (MVA) program in SPSS. Overall, missing data constituted less than 1% of the total. *Little's Missing Completely at Random* test undertaken on this dataset was non-significant ($p=.967$) indicating that missing data can be imputed without systematic error. Imputation via MVA was used to replace the missing data. The means and standard deviations of student's stanine scores in each section of the English GRADE at time 1 for the imputed data was ($M= 4.11, SD=1.72$) for English listening comprehension and ($M=4.95, SD=1.63$) for reading comprehension. Time 2 results were ($M=4.83, SD=1.67$) for English listening comprehension and ($M=5.29, SD= 1.83$) for reading comprehension suggesting that our sample was broadly of typical ability compared to (monolingual) standardisation norms for the GRADE developed in the United States. The means and standard deviations of raw scores for all listening and reading comprehension tests are displayed in Table 1. Inspection of Table 1 shows signs of post-test advantage on all variables and Cohen's d reveals a medium effect size for growth in all measures from pre- to post-test. Effect sizes of around .3-.4 are typical or expected levels of growth in attainment (e.g. Hattie, 2009). Subsequent, t -test analyses run to confirm that all of the scores on the English (L1) measures showed these were all significantly higher than the corresponding L2 measures. This proved to be consistently the case ($p < .05$ in all analyses).

Inferential Analyses on Main Research Questions

Our main research questions relate to the overall similarity and the stability of AIMS teacher practice sub-scores in English and French and over time between fall and winter semester observations. The means, standard deviations and *t*-values are also reported for the AIMS questionnaire, which measures classroom-level variables, in Table 2. Inspection of Table 2 reveals moderate increases in both English and French classroom means in each AIMS category (i.e. Atmosphere, Instruction, Management and Student Engagement) from time 1 to time 2. The *t*-value indicates whether a true significant difference is found between the English and the French classroom, only one significant difference is revealed in the Instruction sub-scale between English and French teachers in the Fall semester at time 1 ($p < .05$). Lower scores of the French teachers at this time reflect lower scores on the components of the Instruction sub-scale: namely lower observed differentiation, instructional density, modeling of strategies, and lower monitoring and support.

To establish the similarity of AIMS sub-scores over time between fall and winter semester observations, correlation analyses were run for all 4 AIMS subtests in English and French across semesters. Correlations exploring each AIMS category in the English and French classrooms to establish the stability of variables over time are reported in Table 3. Table 3 reveals significant associations between English AIMS scores at time 1 and time 2, more notably in the Atmosphere and Classroom Management sub-scales. For the French classrooms however, the only significant association was for the Atmosphere sub-scale.

A second research question concerns the extent to which AIMS measures predicted both current attainment and growth in attainment both in English and French. Given that children in

this sample are nested within classrooms, hierarchical linear modeling (HLM) was used to explore both individual-level and classroom-level variation in students' development of listening and reading comprehension skills in relation to the classroom effects of their English and French classroom. It was first important to establish that significant variability existed between classrooms with respect to student listening and reading comprehension means to warrant hierarchical linear modeling. This was explored by analysis of the unconditional model. The results indicated that the intercepts were statistically significant for all dependent variables, meaning that there is a statistically significant degree of variability between classrooms on each of the listening and reading comprehension outcomes ($p < .001$ in all cases). Inspection of intra-class correlation coefficients indicated that the proportion of classroom-level variance in comprehension measures at post-test was as follows: English Listening Comprehension (8%), English Reading Comprehension (8%), French Listening Comprehension (14%), French Reading Comprehension, (15%). Since the unconditional model demonstrates that there exists significant variability both within and between classrooms, both student- and classroom-level predictors were added to subsequent models to attempt to explain this variability.

A two-level hierarchical model that enabled for the interaction of student-level (level 1) and classroom-level variance (level 2) was conducted (see e.g. Deault & Savage, 2013; Savage et al., 2013 for detailed description of this process and of preliminary model building). Both Level 1 (i.e., individual-level predictors, pre-test scores) and Level 2 (i.e., classroom-level predictors, AIMS factors) were added sequentially to the model to compare the results at different levels of analysis. At Level 1, the student-level predictor was entered into the One-way ANCOVA Model with Random Effects: (a) pre-test variable, to explore the relationship to the dependent variable. For each dependent variable, the pre-test covariate was retained following the results of the

ANCOVA analysis. At Level 2, the Means-as-Outcomes Model, averages for the four AIMS classroom-level Atmosphere, Instruction, Management, and Student Engagement predictors were each added to the model as fixed effects.

Results of analyses revealed significant effects for all four sub-scales of the AIMS at time 1 and time 2 for English Listening Comprehension. For A I M and S respectively at time 1: $F(1, 16) = 2.10, p < .01$, $F(1, 16) = 2.68, p < .001$, $F(1, 6) = 2.47, p < .05$, $F(1, 5) = 2.79, p < .05$, and at time 2: $F(1, 15) = 2.78, p < .001$, $F(1, 14) = 2.94, p < .001$, $F(1, 8) = 3.61, p < .001$, $F(1, 5) = 4.48, p < .01$. Results of analyses revealed significant effects for all four sub-scales of the AIMS at time 1 and time 2 for English Reading, with the exception of time 1 student engagement. For A I M and S respectively at time 1: $F(1, 16) = 2.84, p < .01$, $F(1, 15) = 2.56, p < .001$, $F(1, 6) = 2.73, p < .001$, $F(1, 5) = 1.56, n.s.$, and at time 2: $F(1, 15) = 2.53, p < .001$, $F(1, 14) = 2.97, p < .001$, $F(1, 8) = 2.93, p < .001$, $F(1, 5) = 4.66, p < .001$.

Results of analyses also revealed significant effects for all four sub-scales of the AIMS at time 1 and time 2 for French Listening comprehension. For A I M and S respectively at time 1: $F(1, 16) = 4.24, p < .001$, $F(1, 15) = 4.79, p < .001$, $F(1, 11) = 3.93, p < .001$, $F(1, 7) = 5.96, p < .001$, and at time 2: $F(1, 15) = 4.68, p < .001$, $F(1, 15) = 3.94, p < .001$, $F(1, 9) = 3.95, p < .001$, $F(1, 4) = 4.22 p < .01$. Lastly, results of analyses revealed significant effects for all four sub-scales of the AIMS at time 1 and time 2 for French Reading. For A I M and S respectively at time 1: $F(1, 17) = 4.38, p < .01$, $F(1, 15) = 4.95, p < .001$, $F(1, 11) = 4.75, p < .001$, $F(1, 7) = 3.42, p < .01$, and at time 2: $F(1, 15) = 4.17, p < .001$, $F(1, 15) = 4.32, p < .001$, $F(1, 9) = 3.05, p < .05$, $F(1, 4) = 5.05, p < .01$. In all analyses undertaken in either language, higher AIMS scores were associated with higher growth in Time 2 attainment.

Finally we explored whether there was a relationship between AIMS in English and French for each of the 18 pairs of English and French teachers of the same cohorts of children. All correlations are reported in Table 4, with the key correlations of interest for A I M and S in English and French reported in bold. Inspection of Table 4 reveals that none of these correlations was significant. This suggests that there is little or no consistent pattern of observed AIMS teaching process ratings within a given school for pairs of English or French teachers.

Discussion

The present study sought primarily to explore whether overall classroom teaching quality ratings for literacy instruction using the AIMS (Roehrig & Christesen, 2010), were similar in a cohort of teachers in English and French classrooms serving the same students. We tested claims of poorer quality ‘transmission’ methods of teaching French (Cummins, 1998, 2007a); implicit theories concerning the universality of aspects of quality teaching as ‘general instructional expertise’ (Cummins, 2007b; Pressley et al., 2001); and of ‘goodness-of-fit’ models (Deault & Savage, 2013) in French (generally) an L2 here. In undertaking this work we focused specifically on literacy instruction and in doing so we also assessed the psychometric reliability and validity of AIMS in an L2 context.

Results of group comparisons on AIMS ratings across 18 English and French classrooms revealed only one significant difference in eight contrasts - for the Instruction sub-scale in the Fall semester. Here, lower scores were evident for the French teachers compared to the English. Overall though, the AIMS ratings are not significantly different across languages and such results are thus not consistent with claims that French language teaching uses a more transmission-based approach than English (Cummins, 1998), at least as it applies to the comprehension measures we

used here. Such a transmission model predicts lower Atmosphere and Management, and possibly, higher Instruction scores in French over English classrooms here. The finding that the French teachers scored lower on the ‘I’ of AIMS than their English counterparts in the fall only for the group as a whole is also tempered by inspection of individual teacher scores that revealed that one French second language teacher had the highest Instruction sub-scale scores of the whole sample, exceeding all their English first language colleagues. This suggests there is nothing intrinsic in the literacy curriculum, the children themselves, or the task of teaching L2 literacy that prohibits the practice of highly constructivist approaches per se.

Findings of generally similar AIMS ratings do not provide support for one version of goodness-of-fit models (Connor 2009, 2011; Deault & Savage, 2013). In such models, greater observed independence and self-regulation in classrooms might be associated with greater growth in more capable children. Given that by definition, levels of language attainment and experience are generally higher in our primarily L1 English than our (primarily) L2 French contexts it was reasoned that the goodness-of-fit perspective predicts that respective teachers might adopt seemingly ‘optimal’ literacy teaching. In English classrooms this would predict higher observed Atmosphere and Management scores and lower Instruction scores, with the reverse pattern evident in French. This was not found. One reason for this might be that goodness-of-fit instructional is harder to detect in older children as they have a measure of independence befitting their chronological age, and due to their being more fluent in reading words in texts. The original studies were undertaken in grade 1 classrooms (e.g. Deault & Savage 2013), with beginner readers. It is also possible that at this age children needing more tailored support receive it in other contexts such as resource rooms.

Another major goal in evaluating claims about general instructional expertise was to assess the predictive validity of AIMS constructs in English and in French comprehension. Overall, each of these A, I, M and S measures was a strong predictor of both concurrent attainment and growth in reading and listening comprehension over the course of grade 4. Our analyses here were relatively sophisticated, using hierarchical ‘nested’ analyses of the shared variance in whole classrooms. These results thus cannot reflect the effects of a small number of ‘extreme cases’ of individual children. In the longitudinal analyses, we also controlled for ability at the beginning of the year in reading and listening respectively in all analyses, so the predictive power of these analyses reflects the relative growth made in year 4, not of ability or experience up to year 4. These findings suggest that irrespective of language level, that the same general features of observed teaching predict growth. Theoretically these findings are consistent with the existence of ‘universals’ in the sense of broad professional similarities in quality teaching that are relatively independent of the language of instructional and the level of language experience, at least within regular grade 4 English-French classrooms (Cummins, 2007b; Pressley et al., 2001).

It is crucial to note here that the present results do not suggest that English L1 and French L2 comprehension be taught identically, or provide a reductionist view that ‘teaching is teaching’, or in way devalue the unique complex pedagogical skills needed in French, and as an L2. Rather we test the more modest claim that there exists some significant shared ‘general instructional expertise’ in English and French reliably assessed using AIMS. The finding of some overlapping general instructional expertise is perhaps predictable because influential models of excellence developed for additional languages share content with those from primarily monolingual contexts. For example, Cummins (2014) Literacy Engagement Framework identifies the need to scaffold meaning, connect to student lives, active prior knowledge, affirm identity, and extend language

use beyond the classroom. Cummins (2007a) emphasizes student motivation. Mansour and Butler-Kisber, (2015) note the importance of planning authentic classroom conversations and intentionally teaching vocabulary and other word-learning strategies to unlock meaning. All of these pedagogical themes are arguably evident in research on effective teaching underpinning AIMS (Pressley al., 2001).

Our final question concerned how this association played out on a school-by-school basis. Results here showed no significant correlation between any of the English AIMS and French AIMS sub-tests across the 36 classrooms observed. This suggests that while across the sample the same features (high scores on AIMS) predict outcomes in both E and F, and that overall, AIMS scores did not vary by language, children in any given school generally moved between English and French language classrooms with quite different teaching styles. These results might suggest relatively limited collaboration between teachers within schools. Attempts to improve this situation and their effectiveness were described earlier (e.g. Gunning et al., 2016; Lyster et al., 2013), and the need to do so is clearly pressing. Our findings here are nuanced by the observation that as often as not it was French (generally L2), teachers who showed greater pedagogical excellence, best practices that could be shared with English teacher partners.

Limitations in the present study

It is important to note that the classes were not L1 and L2 respectively for all of these children – two thirds of the sample parents reported English as L1 with 16% reporting French as an L1. The present results with linguistically diverse cohorts reflected the complex patterns evident on the ground in our sample and the contexts where tools such as AIMS show (and must show) reliability and validity. Our results may reflect certain particularities of local instructional

practices and certainly require replication. We were unable to administer a test of word reading in this study. There were sometimes low reliabilities for certain tests, namely listening comprehension in French and English. Modest reliabilities for listening comprehension sub-test of the GRADE have been reported in some other studies (e.g. Savage et al., 2013) and is a problem common to the whole literature on bilingual and dual language instruction (e.g. August & Shanahan, 2006). There was however some consistency (cross semester test-retest reliability) for AIMS scores for Atmosphere and Management subscales for English and Management for French. We also had relatively limited information about the breakdown teachers from each school and the nature of their pairings. This occurred because a condition of the study that we obtain only the broadest data form the teacher participants and that we report no identifiable data on their characteristics. While each of our partner school boards espoused a committed policy of French immersion, there may be school-board level effects that we are not able to elucidate or directly analyse. It should be noted however that schools in these boards have, through their governing bodies, jurisdiction over policy, and the broadest variations in practice we observed existed within a school board, rather than between them.

Implications and Future Research

The present paper reported theoretically driven analyses using the AIMS observational tool, showing that there was a quite similar pattern of practice that best predicted growth in attainment, a ‘universal’ of global features of effective teaching in both L1 and L2 contexts. Theoretically this might set the stage for future work developing tools assessing the unique aspects of immersion-specific pedagogy such as use of dual language multimedia (Lyster et al. 2009), explicit teacher cross-reference from L1 to L2 (Cummins, 2014; Horst, White, & Bell,

2010) amongst many other L2-specific techniques (Horwitz, 2012). Here we hope that AIMS might serve as a preliminary benchmark for the development of nuanced future measures.

Practically our results suggest that AIMS can be used to assess the quality of French (generally L2) teaching. Valid assessments of teaching quality might be used to audit and support French (L2) teaching in experienced and novice teachers and might also facilitate development of communities of practice (Bannister, 2015). They can be used to assess teacher change following professional development. Assessments of teaching are an important and sometimes overlooked part of comprehensive assessments of student's needs, particularly for high-stakes decision making such as school or curriculum placement. Such assessments can be quite within child- and deficit- focused. Some contemporary models such as RtI emphasize the centrality of also assessing instructional quality, including instruction in L2 (Genesee et al., 2013; Jimerson, et al., 2016). The present work provides part of a technology for doing so, demonstrated in French majority L2 contexts.

Declaration of Interest Statement

The authors have no relevant conflicts of interest of any kind in relation to this work.

References

- August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth.* Mahwah, NJ: Lawrence Erlbaum Associates.
- Bannister, N. A. (2015). Reframing practice: Teacher learning through interactions in a collaborative group. *Journal of the Learning Sciences*, 24, 347-372. doi: 10.1080/10508406.2014.999196
- Connor, C., Morrison, F. J., Fishman, B., Ponitz, C.C., Glasney, S., Underwood, P., et al., (2009). The ISI classroom observation system: Examining the literacy instruction provided to individual students. *Educational Researcher*, 38, 85-99. doi: 10.3102/0013189x09332373
- Connor, C., Morrison, F. J., Schatschneider, C., Toste, J., Lundblom, E., Crowe, E. C et al., (2011). Effective classroom instruction: Implications of child characteristics by reading instruction interactions on first graders word reading achievement. *Journal of Research on Educational Effectiveness*, 4, 173-207. doi: 10.1080/19345747.2010.5101079
- Cummins, J. (1998). Immersion education for the millennium: What have we learned from 30 years of research on second language immersion? In M. R. Childs & R. M. Bostwick (Eds.) *Learning through two languages: Research and practice. Second Katoh Gakuen International Symposium on Immersion and Bilingual Education.* (pp. 34-47). Katoh Gakuen, Japan.

- Cummins, J. (2007a). Pedagogies for the poor? Re-aligning reading instruction for low-income students with scientifically based reading research. *Educational Researcher*, 36, 564–572.
- Cummins, J. (2007b). Rethinking monolingual instructional strategies in multilingual classrooms. *Canadian Journal of Applied Linguistics*, 10, 221–240.
- Cummins, J. (2014). Beyond language: Academic communication and student success. *Linguistics and Education* 26, 145–154
- Deault, L., & Savage, R. S. (2013). Effective classroom contexts to develop literacy and attention skills for typical and atypical children. *Psychological Abnormalities in Children*, 1, 1-9.
doi: <http://dx.doi.org/10.4172/2329-9525.1000106>
- Fugate, M. H. (2003). Review of the Group Reading Assessment and Diagnostic Evaluation. In B. S. Plak, J. C. Impara and R. A. Spies (Eds.), *The fifteenth mental measurements yearbook*. (pp. 425-431). The Buros Institute of Mental Measurements: Lincoln, Nebraska.
- Genesee, F., Geva, E., Dressler, D., & Kamil, M. (2006) Synthesis: Cross-linguistic relationships. In D. August & T. Shanahan (Eds), *Developing literacy in second- language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (pp. 153-174). Mahwah, NJ: Lawrence Erlbaum.
- Genesee, F., & Jared, D. (2008). Literacy development in early French Immersion programs. *Canadian Psychology*, 49, 140-147.
- Genesee, F., Savage, R., Erdos, E., & Haigh, C. (2013). Identification of reading difficulties in students schooled in a second language. In Gathercole, V. (Ed.). *Solutions for the assessment of bilinguals*, Vol 2, (pp. 10-35). Clevedon: Multilingual Matters.

Gunning, P., White, J., & Busque, C. (2016). Raising learners' awareness through L1–L2 teacher collaboration, *Language Awareness*, 25, 72-88. doi: 10.1080/09658416.2015.1122022

Hamayan, E., Genesee, F., & Cloud, N. (2000). *Dual Language Instruction from A to Z*. New York: Benjamin. doi: 10.1075/jicb.2.1.09hoo

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first grade classroom make a difference for children at risk of school failure? *Child Development*, 76, 949-967.

Hattie, J. (2009). *Visible learning*. London: Routledge.

Horst, M., White, J., & Bell, P. (2010). First and second language knowledge in the language classroom. *International Journal of Bilingualism*, 14, 331-349.

Horwitz, E. K. (2012). *Becoming a language teacher: A practical guide to second language learning and teaching* (2nd Edn). Boston: Pearson Education

Jimerson, S. R., Burns, M. K., & van der Heyden, A. M. Eds. (2016) 2nd Edition. *Handbook of Response to Intervention: The Science and Practice of Multi-Tiered Systems of Support*. New York: Springer.

Johnson, K., & Swain, M. (1997). *Immersion Education: International Perspectives*. Cambridge: Cambridge University Press.

Kirby, J., & Savage, R. S. (2008). Can the simple view deal with the complexities of reading? *Literacy*, 42, 75-82 (special issue on the Simple View of Reading).

Lambert, W., & Tucker, R. (1972). *Bilingual education of children: The St. Lambert experiment*. Rowley, MA: Newbury House.

- Louden, W., Rohl, M., Barratt-Pugh, C., Brown, C., Cairney, T., Elderfield, J., ...Rowe, K. (2005). In teachers' hands: Effective literacy teaching practices in the early years of schooling. *Australian Journal of Language and Literacy*, 28, 171-253.
- Lyster , R., Collins, L., and Ballinger, S. (2009). Linking languages through a bilingual read-aloud project. *Language Awareness*, 18, 366–38.
- Lyster, Quiroga, J., & Ballinger, S. (2013). The effects of biliteracy instruction on morphological awareness. *The Journal of Immersion and Content-based Language Instruction*, 1, 169-197.
- Mansour, A., & Butler-Kisber, L., (2015). Collaboration between language teachers and instances of cross-language learning transfer in two languages ... a joint approach. Literature Review Submitted to Ministry of Education, Leisure and Sports (MELS) of Quebec Office of Leadership in Community and International Initiatives (LCII) McGill University. <http://blogdev.learnquebec.ca/facet/wp-content/uploads/sites/7/2014/02/STEF-literature-review-Report-2014.pdf>
- Melby- Lervåg, M., & Lervåg, A. (2011). Cross- linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta- analysis of the correlational evidence. *Journal of Research in Reading*, 34, 114-135.
- National Reading Panel (2000). Teaching children to read: reports of the subgroups. <http://www.nichd.nih.gov/publications/nrp/report.htm>
- Pianta, R. C., Le Paro, K. M., & Hamre, B. K., (2007). *Classroom Assessment Scoring System- CLASS*. Baltimore: Brookes.

- Pressley, M. (1998). *Reading Instruction That Works*. New York: Guilford.
- Pressley, M., Wharton-McDonald, R., Allington, R., Block, C. C., Morrow, L., Tracey, D., ...
- Woo, D. (2001). A study of effective first-grade literacy instruction. *Scientific Studies of Reading*, 5, 35-58.
- Roehrig, A. D., & Christesen, E. (2010). Development and use of a tool for evaluating teacher effectiveness in grades K-12. In V. J. Shute, & B. J. Becker (Eds.), *Innovative Assessment for the 21st Century* (pp. 207-228). New York, NY: Springer.
- Roehrig, A. D., Bohn, C. M., Turner, J. E., & Pressley, M. (2008). Mentoring beginning primary teachers for exemplary teaching practices. *Teaching and Teacher Education*, 24, 684-702.
- Savage, R. S., Abrami, P., Piquette-Tomei, N., Wood, E., Sanghera-Sidhu, S., Burgos, G., & Deleveaux, G. (2013). A pan-Canadian cluster randomized control effectiveness trial of the ABRACADABRA web-based literacy program. *Journal of Educational Psychology*, 105, 310-328. doi: 10.1037/a0031025.
- Savage, R. S., & Côte, M-F. (2017). The teaching of reading and writing in second and multi-language contexts. In A Boulos (Ed.), *Teaching and Learning Literacy in the 21st Century: Lessons from Around the World*. Paris: IBE/UNESCO
- Savage, R. S., & Cloutier, E. (2017). Early reading interventions: The state of the practice, and some new directions in building causal theoretical models. In Cain, K., Compton, D., and Parrila, R. (Eds.), *Theories of reading development*. (pp. 409-436).

Savage, R. S., & Pace, A. (2017, in press). Linguistic and reading comprehension in simultaneous dual language instruction: Evidence against unitary constructs. *International Journal of Bilingualism*.

Tabachnick, B.G. & Fidell, L.S. (2007). *Using multivariate statistics*. (4th edn). New York: Allyn & Bacon.

Topping, K., & Ferguson, N. (2005). Effective literacy teaching behaviours. *Journal of Research in Reading*, 28(2), 125-143.

Waterman, B. B. (2003). Review of the Group Reading Assessment and Diagnostic Evaluation. In B. S. Plak, J. C. Impara, & R. A. Spies (Eds.), *The fifteenth mental measurements yearbook*. (pp.425-431). Lincoln, NE: The Buros Institute of Mental Measurements.

Williams, K. T. (2001). *Group Reading Assessment and Diagnostic Evaluation: Teacher's scoring and interpretive manual*. Circle Pines, MN: American Guidance Service.

Appendix: Exemplification of AIMS content

Atmosphere (2 specific example item in brackets for each sub-category)

Sense of community fostered (teacher expresses empathy/ value students; altruism is nurtured)

Democratic classroom established (it is a cooperative environment; the classroom is a democratic place)

Interest fostered (teacher builds anticipation; teacher encourages curiosity/suspense)

Focus on effort rather than performance (teacher encourages a changeable concept of intelligence; teacher attributes success to effort and time)

High expectations expressed (teacher communicates high expectations; that all students can and will learn)

Informative feedback provided (teachers cues students to the positive /partly correct; teacher does not grade publicly)

Instruction

Engaging content and activities used (tasks match to students interests; students lived experience integrated)

Instruction individualized- high instructional density achieved (classroom is busy; mini-lessons)

Instruction well-planned- high instructional density achieved (students never with nothing to do; activities well-organized)

Appropriate challenge level achieved (teacher supports risk-taking; teacher provides appropriately challenging materials)

Thinking process modeled and taught (teacher explicitly models strategies; encourages higher-order thinking)

Management

Behavioral self-regulation encouraged (teacher establishes routines that are automatized; teacher expresses expectation of student self-regulation)

Behavior/Task monitoring provided (teacher does not scapegoat; uses no public punishment)

Student Engagement

At least 80% of students are consistently on task; students vocalize excitement (lots of oohs, aahs)

Table 1

Means, Standard Deviations and Proportions of GRADE Comprehension Measures

Measure	Pre-test				Post-test				<i>d</i>
	M	SD	Max	%	M	SD	Max	%	
English Listening	13.66	1.90	17	.80	14.58	1.74	17	.86	.51
English Reading	16.68	5.45	28	.60	18.92	5.78	28	.68	.40
French Listening	13.15	2.00	17	.77	13.93	1.97	17	.82	.39
French Reading	14.65	5.40	28	.52	16.44	5.90	28	.59	.32

Note. All measures are raw scores. % = M/Max

Table 2

Means, Standard Deviations and *t*-tests of AIMS measure

	Time1 (Fall semester)			Time 2 (Winter semester)		
	English	French	T	English	French	t
Atmosphere	2.59 (.35)	2.42 (.41)	1.33	2.71 (.29)	2.56 (.33)	1.36
Instruction	2.44 (.25)	2.18 (.38)	2.50*	2.50 (.33)	2.35 (.36)	1.31
Management	2.70 (.47)	2.55 (.37)	1.04	2.73 (.31)	2.64 (.37)	0.73
Engagement	2.60 (.45)	2.41 (.45)	1.28	2.69 (.37)	2.57 (.34)	0.98

Note. * <0.5 ; $df = 17$ in all cases.

Table 3

Correlations between AIMS (English) at Time 1 and Time 2

A) English measures		T2 English Atmosphere	T2 English Instruction	T2 English Management	T2 English Student Engagement
T1 English Atmosphere		.65**	.47	.31	.42
T1 English Instruction		.48*	.33	.18	.37
T1 English Management		.65**	.49*	.68**	.68**
T1 English Engagement		.46	.27	.36	.32
B) French measures		T2 French Atmosphere	T2 French Instruction	T2 French Management	T2 French Student Engagement
T1 French Atmosphere		.55*	.44	.13	.41
T1 French Instruction		.29	.26	.21	.22
T1 French Management		.37	.12	.34	.28
T1 French Engagement		.44	.10	.51*	.33

Note. *p < .05, **p < .01, ***p < .001.

Table 4

Correlations between AIMS Measures at Time 1 and Time 2 in English (L1) and French (L2).

Measure	1	2	3	4	5	6	7	8
1 English Atmosphere	-	.67**	.61**	.69**	.07	-.04	.05	-.20
2 English Instruction	.83**	-	.29	.53*	-.24	-.31	-.03	-.13
3 English Management	.47*	.37	-	.67**	.08	-.12	-.13	-.18
4 English Engagement	.84**	.70**	.67**	-	-.17	-.44	-.15	-.38
5 French Atmosphere	-.02	-.06	.09	-.01	-	.75**	.61**	.57*
6 French Instruction	.01	.07	-.18	-.22	.78**	-	.58*	.76**
7 French Management	-.01	-.12	.21	.15	.74**	.45	-	.74**
8 French Engagement	.08	-.08	.38	.18	.76**	.52*	.72**	-

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Upper right diagonal: correlations between ELA and FSL teachers at time 1; Bottom left diagonal: correlations between English and French teachers at time 2

Figure 1. Flow of Participants in the Present Study

