

Accepted draft for publication:

INTERNATIONAL JOURNAL OF BILINGUAL EDUCATION AND BILINGUALISM, 2017

- <https://doi.org/10.1080/13670050.2017.1407289>

To cite this article: Josep Coral, Meritxell Urbiola, Eduard Sabaté, Josep Bofill, Teresa Lleixà & Ruth Vilà Baños (2017): Does the teaching of physical education in a foreign language jeopardise children's physical activity time? A pilot study, International Journal of Bilingual Education and Bilingualism, DOI: 10.1080/13670050.2017.1407289

Does the teaching of physical education in a foreign language jeopardise children's physical activity time? A pilot study.

Coral, Josep, Meritxell Urbiola, Eduard Sabaté, Josep Bofill, Teresa Lleixà and Ruth Vilà

Abstract

The purpose of this study was to design and validate a tool to observe the teaching of physical education (PE) through a foreign language (L2) and to measure the engaged time and language learning aspects of children's physical activity in PE classes where a content and language integrated approach (CLIL) was used. A first draft of the tool was tested for validity by experts, revised and then tested in a pilot study by three trained observers to 613 minutes of video-recordings of 15 different CLIL teachers conducting real PE classes. The results of the pilot study indicated that motor-engaged time in PE lessons applying the CLIL approach was, on average, 41.86%, which is lower than what is generally recommended by educational authorities for PE classes. Non-parametric tests indicated that this shortfall in the amount of physical activity time may be accounted for by the excessive use of language support materials such as flashcards to facilitate communication or the types of physical activity involved in the lessons. It is suggested that this imbalance in physical activity time vs. L2 language learning could be addressed by either extended the duration of PE-in-CLIL classes or by providing more focused training for PE-in-CLIL teachers.

Key words:

CLIL

Foreign language
Physical Education,
Students' activity time,
Validation procedures.

1. Introduction

An increasing number of schools around Europe offer content and language integrated learning (CLIL) programmes (Coyle, Hood and Marsh 2010; Dalton-Puffer 2011; Lorenzo, Casal and Moore 2011) and physical education (PE) is one of the primary subject areas (Coral 2013; Devos 2012; Martínez-Hita and Garcia-Canto 2017; Rottmann 2006). Little research, however, has addressed the effectiveness of the CLIL approach in PE, raising questions about how CLIL theory affects the achievement of PE contents, and more specifically, whether teaching PE using a CLIL approach may affect the amount of time that children are involved in physical activity. Thus, the purpose of this study was to design and validate a tool to observe the teaching of physical education (PE) through a foreign language (L2) and to measure the engaged time and language aspects of children's physical activity in PE classes in which a content and language integrated approach (PE-in-CLIL) was used. The results are part of a wider action research (AR) study of a teacher training programme (Coral and Lleixà 2017) carried out with the support of the *Centre de Recursos Pedagògics Específics de Suport a la Innovació i la Recerca Educativa* (Centre for Teaching Resources for Innovation and Research in Education, hereafter CESIRE) of the Ministry of Education of the Government of Catalonia.

1.1 Students' engaged time in physical activity

The time that students are engaged in physical activity has for years been considered an indicator of effective PE teaching (Pieron 1999, Siedentop et al. 1982, Silverman 1985). Thus, at the end of the last century, several studies looked at time in PE and sports pedagogy. At that point, the main areas of study were how time was conceptualised and measured in physical education and sports settings, and how time spent in physical education related to the way that students learn (Metzler 1989).

The concept ALT-PE (academic learning time-physical education) arose in this context and was applied by Siedentop et al. (1982) in their codification of time in PE classes. The ALT-PE tool had an enormous influence on the other observational tools focusing on time aspects. This was noted by Lozano et al. (2006), who conducted an analysis of observational instruments for the study of time variables in physical education settings. Nowadays, the SOFIT tool (system for observing fitness instruction) (McKenzie et al. 1991) is widely used because it is easily available and measures other features related to teaching. Moreover, Weaver (2016) presents a modified version called SOFIT+, which comprehensively measures best practices to promote moderate-to-vigorous physical activity (MVPA). Other less frequently used tools include the accelerometer, heart rate monitoring, and the Beaches or continuous direct observation method (Hollis et al. 2015).

The time students are involved in motor activity has been used as a variable in many studies carried out in different contexts and with different research goals. We can find research in preschool (Chow et al. 2015), primary school (Smith et al. 2015; Hollis et al. 2015) and secondary school (Viciano et al. 2012). The objectives are mainly concerned with skills acquisition (Derri et al. 2007), concept acquisition (Derri et al. 2008), teaching strategies (Yanci et al. 2016; Viciano et al. 2012) and teacher behaviour (Behets 1997; Senne et al. 2010).

More recently there have been a large number of new studies on the subject arising from the fact that many international organisations recommend increasing the time devoted to physical activity and, particularly, MVPA in children. These studies tend to analyse the effectiveness of interventions and methodological approaches to increase the time spent by students in MVPA (Bevans et al. 2010; Olmedo 2000). For example, Harvey et al. (2016) measure students' PA levels over a PE football unit designed using a game-centred approach. Logan et al. (2015) observed that students engaged in significantly more MVPA in mastery and performance climates than in typical PE. Recently, and more specifically related to our research topic, Martínez-Hita and Garcia-Canto (2017) analysed the relationship between the introduction of bilingual CLIL approaches in PE lessons and children's motor-engaged time. They found significant differences in the time spent on the task between monolingual and bilingual schools, with bilingual schools spending less time.

1.2 Teaching PE in Content and Language Integrated Learning: PE-in-CLIL

In Europe, the emergence of English as a lingua franca has had a significant impact and the CLIL approach has been gaining in popularity, influencing ways of teaching content subjects through a foreign language, mainly English. According to Coyle et al. (2014, 4) CLIL is ‘neither language learning nor subject learning, but an amalgam of both’ while Gajo (2007, 563) describes CLIL as ‘a powerful means of teaching a subject through a second language, thus enhancing the latter by the means of teaching and learning the former’. In the current competence-based curriculum, CLIL can be linked to a process of convergence between the teaching and learning of both content and language taking into account knowledge, skills and attitudes for learning. Coral and Lleixà (2016) understand PE-in-CLIL as a holistic approach that uses the principles of learning-by-doing to teach motor contents through the medium of a foreign language using CLIL approach. In this context, teaching PE-in-CLIL should be regarded as an opportunity both to find new ways of teaching PE and to enhance communication in a foreign language. It also takes into consideration pupils’ motivation for movement and provides language support to achieve both motor and language goals. This is consistent with the 4Cs CLIL framework laid out by Coyle et al. (2010)¹ as it not only provides learners with motor skills, healthy habits and interaction in a foreign language, but also contributes to the acquisition of learning strategies, thinking skills and cultural awareness. Additionally, when the 4Cs are applied through PE, civic and social competences must be also considered since in a competence-based curriculum PE aims to integrate physical activity in daily life by giving value to its role in mental, emotional, social and physical well-being and civic coexistence (Generalitat de Catalunya 2015, 2016).

CLIL is usually applied in subjects other than language in parallel with mainstream foreign language lessons, and improvements in student L2 performance can be attributed to both the CLIL approach and the increased length of exposure to the language. With regard to learning outcomes, some authors argue that CLIL programmes are more successful in developing foreign languages than traditional language classes are (Dalton-Puffer 2011; Ioannou-Georgiou 2012; Lasagabaster 2008; Navés 2009;

¹ The 4Cs stands for Content, Communication, Cognition, Culture

Lorenzo et al. 2009). In terms of language proficiency, Admiral et al. (2006) in a longitudinal CLIL study in secondary education in the Netherlands showed higher scores in oral proficiency and reading comprehension when compared to a non-CLIL group control. More recently, Ouazizi (2016, 129) concluded that ‘CLIL education leads to a better language proficiency in the target language because it seems to activate implicit mechanisms of learning a language as displayed in students’ simultaneous learning and practicing the target language’. Similarly, research into the levels of content achievement also shows positive results (Bonnet 2012), as another longitudinal study in Belgium (Surmont et al. 2016) demonstrates by providing evidence that CLIL apparently has a positive impact on students’ mathematical performance, thus confirming that the acquisition of subject matter can be enhanced by a CLIL environment (Van de Craen et al. 2007). Other studies such as Sylvén (2013) discuss the discrepancies in CLIL outcomes in four European countries, while Bruton (2011) expresses his doubts about the analysis and conclusions of some CLIL research. Related to the context where CLIL is applied, on the one hand we agree with Paran (2013) that there are two factors which are crucial for a CLIL programme to be successful: a) high level of investment in teachers and teaching skills in general, and b) a high level of L2 mastery on the part of CLIL teachers. On the other hand, we disagree with the idea that CLIL is related to practices of student selection such that only students who are already ‘high achievers both in language and content’ should be allowed to participate in CLIL programmes (Paran 2013, 327), not least because this contradicts the evidence provided by the present study, which involved public schools that form part of a comprehensive school-system with heterogeneous and inclusive class groups. Overall, we are in full agreement with the point made by Coyle (2007) and Cenoz, Genesee and Gorter (2013) about the need to conduct more research that is generalizable, meaningful and useful using a wide range of research opportunities and approaches.

Most of the specific PE-in-CLIL studies, although scarce, are focused on L2 achievement, and their findings also point to an improvement in students’ mastery of the target language (Coral and Lleixà 2014; Coral, Lleixà and Ventura 2016; Devos 2016; Rottmann 2006). A few are related to other aspects such motivation. For example, Heras and Lasagabaster (2015) claim that PE-in-CLIL can help reduce gender differences in motivation and has a similar positive effect in both genders on the acquisition of technical content-related vocabulary. A recent study is centred on content

achievement (Fernández-Sanjurjo et al. 2017) and shows that CLIL students in primary school slightly underperform in their acquisition of science contents when assessed in their mother tongue. However, to our knowledge there is no evidence that teaching PE-in-CLIL has a negative effect on children's achievement in the subject or that teaching language-related elements somehow reduces children's engagement in physical activity time. This is especially important since teaching PE-in-CLIL usually involves language techniques to enhance oral interaction, and the use of special support materials to foster both comprehension and production could conceivably alter the delicate balance between content and language to the benefit of the latter but at the expense of the former.

Recent studies confirm that PE is seen as an opportunity to engage students in healthy physical activity (Erfle and Gamble 2014; Heidorn et al. 2016; Pate et al. 2006). Based on the findings of a systematic review of the subject, Janssen and LeBlanc (2010, 11) recommended that 'children and youth 5-17 years of age should accumulate an average of at least 60 minutes per day and up to several hours of at least moderate intensity physical activity. Some of the health benefits can be achieved through an average of 30 minutes per day'. The World Health Organization (2010) has corroborated these recommendations. Taking into consideration that health is one of the main objectives included in PE curriculums worldwide (see, among others, Australian Curriculum, Assessment and Reporting Authority 2012; Government of Catalonia 2016; Scottish Government 2004), it would seem counterproductive to promote foreign language learning in PE classes if that were to put physical activity time at risk. This crucial question has received very little attention so far in the literature, and in order to build up the body of evidence required to allow generalizable conclusions, the first step would seem to be the creation of a validated tool by which to measure these two factors, language learning and physical activity time, and then compare them across PE-in-CLIL programmes.

Several recent studies involve the design and validation of tools to assess different aspects of PE teaching. Examples include: an instrument specifically devised for the systematic observation of the teaching of games in PE using a five-stage system (Roberts and Fairclough 2012); a Likert scale questionnaire intended to determine how specialist teachers in primary and secondary education add key competences to their PE

teaching programmes (Lleixà et al. 2015); and a questionnaire to measure athletes' perception of the influence of personal characteristics, trainers, training, family and environment on their sports results using a three-step process (Simón et al. 2017). Taken as a whole, these studies show that a wide range of validation instruments are available. Nevertheless, specific areas related to PE teaching still lack validated tools and our goal was to fill one such gap by means of the present study.

1.3 Objective and research questions

The objective of this study is to design and validate a tool to observe PE-in-CLIL lessons and measure not only pupils' engaged time in physical activity but also the language-oriented aspects that are part of the teaching of contents through a foreign language. In particular, the study will explore the following research questions (RQ):

RQ1: What variables and indicators can be used to observe and analyse PE-in-CLIL lessons?

RQ 2: Considering common validation procedures, what procedure can be applied in the context of this study?

RQ 3: Once applied, does the proposed procedure confirm the reliability and validity of the PE-in-CLIL observation tool presented here?

RQ 4: Regarding pupils' engaged time in physical activity, do the results obtained in pilot testing suggest any differences with the main studies related to this topic?

2. Method

3.

2.1 Context

This study is part of an action research (AR) project carried out during the 2014-2016 school years with the support of the CESIRE resource centre and the Faculty of Education of the Autonomous University of Barcelona. The goal of the AR project was to provide teachers with effective, evidence-based and inquiry-based knowledge in order to negotiate CLIL lessons successfully at their schools (Coral and Lleixà 2017). Like all educational AR projects, it sought improvement in the practice of teaching

through reflection and research. One of its aims was to closely monitor research in teaching and education coming from schools, universities and other institutions so that the results could be promoted and adapted to meet teachers' needs. The validation tool presented in this paper is one of the project's outcomes.

2.2 Development process of the observation tool for PE-in-CLIL lessons

A four step model was used to develop the observation tool. These four steps were defining the dimensions, variables and indicators; training observers; pilot testing; and carrying out a descriptive analysis of preliminary data.

2.2.1 Step one: Defining the dimensions, variables and indicators

According to Quivy and Van Carnpenhoudt (2005), complex concepts need to be decomposed into a set of dimensions, variables and indicators, whereby a broad dimension (e.g. 'Activity') is broken into component variables (e.g. 'Motor locomotion', 'Language skills') and then rubrics or 'indicators' describe the specific behaviours that evidence the variable in question (e.g. 'Class time when students were engaged in performing a game-related motor response'). To define the dimensions, variables and indicators (Table 1) that would index the children's activity type in PE-in-CLIL lessons, we followed the behavioural classification used by Roberts and Fairclough (2012) with some adaptations to simplify it and to include language-related aspects that are part of teaching PE-in-CLIL (Coral 2013). Thus, the fundamental division was between two dimensions, 'Activity' and 'Inactivity', which were broken down into three variables ('Listening', 'Oral Interaction' and 'Other') and two variables ('Motor locomotion only' and 'Motor locomotion + L2 language skills'), respectively. All the indicators characterising each variable were chosen on the basis of three factors: first, whether they corresponded to activities that were to be used in PE-in-CLIL lessons; second, whether they were observable; and third, whether they could provide meaningful data relevant to the target variables and dimensions.

A preliminary draft of the tool was submitted for evaluation to a six-member panel consisting of two experts in PE, two experts in both PE and the CLIL approach and two experts in CLIL and language teaching (in this case, English as a foreign language), all of them experienced teacher trainers then revised in accordance with their suggestions.

This revised version (see table 1) provided fuller descriptions of the indicators related to the presence of oracy, particularly the variable ‘oral interaction’. A clearer distinction was also drawn between the two variables of the dimension ‘activity’ regarding the presence or absence of a purposeful speaking action (Coral and Lleixà 2016).

[table 1 near here]

It was felt that information about classroom activity would not be fully meaningful without data characterising the particular context where the lesson was taking place. For this reason, a supplementary checklist of context variables (see table 2). Such context variables included not only general matters like educational level of the class being observed (i.e. primary or secondary) and type of physical activity being carried out but also specific L2 teaching-related aspects like the use of language support materials to facilitate communication or specific language skill training in the lesson. The presence or absence of language support materials such as flashcards to facilitate communication was included as a variable since such support materials are necessary when higher-order thinking skills are required to complete a task (Meyer 2010). Regarding language skill training, speaking was included because enhancing oracy is one of the explicit goals of PE-in-CLIL (Coral and Lleixà 2016; Davos 2016), and reading and writing because literacy is known to facilitate language acquisition in PE classes in primary and secondary school (Bell and Lorenzi 2004).

The revised version of the observation tool was then re-submitted to the panel of six experts, who again independently rated the relevance of each indicator in the tool using a four-point Likert scale (1 = not at all relevant, 2 = slightly relevant, 3 = moderately relevant and 4 = very relevant). In order to identify the indicators judged insufficiently relevant, inter-judge reliability was first checked by applying the formula² proposed by House et al. (1981, 37-57) and then calculating the kappa statistic. According to House et al. (1981), there is considerable consensus that an average of agreement at or above 70% is necessary in order to show that raters are consistent in their judgements among themselves. We decided to be slightly more demanding, so that when we found less than 75% agreement for any given indicator, further clarification was necessary and the judges were asked to revise their judgements in the hope of obtaining a higher kappa

² $(\text{Number of agreements} / (\text{Number of agreements} + \text{Disagreements})) \times 100 = \% \text{ agreement}$

value. Like most correlation statistics, kappa can range from -1 to +1. We used Altman's interpretation (1991): K values between 0.81 and 1.00 indicated very good agreement, between 0.61 and 0.80 good agreement, and between 0.41 and 0.60 moderate agreement, while values between 0.21 and 0.40 indicated fair agreement and any value below 0.20 was poor. In earlier research measuring agreement among observers, Landis and Koch regarded a range between 0.61 and 0.80 as indicating substantial agreement and any rating above 0.81 almost perfect, pointing out that 'although these divisions are clearly arbitrary, they do provide useful benchmarks for the discussion' (1977, 165).

[Table 2 near here]

2.2.2 Step two: *Observer training*

The training of observers followed the distinction between basic and applied training described by Losada and Manolov (2015). In basic training observers become familiar with a tool by 'making successive observations of increasing complexity' (Losada and Manolov 2015, 341), while in applied training observers make the observation using the definitive category system. Three trainee observers were selected, all of them experienced and trained PE-in-CLIL teachers. In their basic training, they were first instructed in the use of LINCE software. This software, which allows for easy mark-up and analysis of video-recorded materials, is specially designed for analysing physical education activities and has demonstrated a high degree of reliability and validity for that purpose (Gabin et al. 2012). Next, the trainees were asked to individually view, code and rate short recordings related to PE-in-CLIL which were provided to them from a corpus of previously recorded samples to develop their practical skills as observers and become familiar with the variables and indicators stipulated in the observation tool. To increase the complexity of what they had to do, they were then each given the same short (4 min 50 s) recording of a PE-in-CLIL task and were asked to use the observation tool to analyse it, working individually. This was followed by the applied training phase, in which the three observers individually used the tool to analyse a much longer 41-minute recorded segment. Both tasks were followed by an evaluation of inter-observer reliability using Krippendorff's alpha for ratio data (Krippendorff 2004). For this reliability test, the minimum acceptable alpha coefficient should be chosen according to the importance of the conclusions to be drawn from imperfect data. When

the costs of mistaken conclusions are high, the minimum alpha needs to be set high as well. In the absence of knowledge of the risks of drawing false conclusions from unreliable data, social scientists commonly rely on data with reliabilities of $\alpha \geq .800$, consider data with $0.800 > \alpha \geq 0.667$ only to draw tentative conclusions, and discard data whose agreement measures $\alpha < 0.667$.

2.2.3 Step three: Pilot testing

Pilot testing was carried out using a corpus of self-made classroom video-recordings of PE-in-CLIL lessons made by a natural sample of 15 teachers who had taken part successfully in CLIL training courses during 2015 and 2016. Ten of these teachers were involved in primary education and five in secondary education. Prior to their participation, they were informed that they would be part of a study that met the ethics requirements of the Autonomous University of Barcelona, that their recordings would be used exclusively for the purposes of the research and then deleted at the end of the study, and that the names of their schools would not be disclosed. Though they were informed that the aim of the study was to analyse the teaching of PE-in-CLIL, they were asked not to modify or change their teaching behaviours in any way. Moreover, the specific aim of identifying activity and inactivity frames in their lessons was not revealed. Finally, they were requested to obtain compulsory permissions from schools and/or parents as a prerequisite to contributing video materials to the research project.

The 15 participating teachers each video recorded one or two lessons (or part of a lesson) and uploaded them in a compatible format to a shared folder in Google Drive. Long to medium-length recordings (30-45 mins) were requested since the objective was to capture the general flow of the lessons, which would allow the observer to identify easily whether the teaching produced activity or inactivity in the children. Also, in order to facilitate the identification of the variables related to language use, the recordings needed to have high audio quality. Pre-lesson and post-lesson routines were not included in the recordings and teachers could record either their full lesson (introduction, main body and conclusion) or a part of it. The result was a corpus of 18 video-recorded segments from roughly 30 to 50 minutes in duration, representing a total of 613 minutes of video footage to be coded using the new tool.

Testing of the observation tool was then carried out by the three trained observers, who coded all recorded segments using LINCE, labelling all dimensions and variables and marking durations in seconds.

2.2.4 Step four: Descriptive analysis of preliminary data

In this step, first the three sets of tool-based data from the observers were checked for inter-observer reliability. That reliability having been found satisfactory, the three sets of data derived from the tool regarding class activities and durations as well as contextual variables were analysed using IBM SPSS Statistics 23.0 in order to obtain descriptive results showing minimum and maximum values expressed in seconds, percentages of time, means and standard deviations. Next, the non-parametric Mann-Whitney U test was used to assess significant differences among three of the contextual variables (educational level, use of language support to facilitate communication, and writing), while the remaining contextual variables (type of physical activity, speaking, and reading) were analysed using the Kruskal-Wallis non-parametric test.

3. Results

3.1. Inter-judge agreement

The results of the inter-judge agreement analysis using the formula proposed by House et al. (1981) and kappa appear in table 3. For the first draft of the observation tool, the kappa values had an overall agreement of 0.56, which corresponds to moderate agreement. When the analysis was applied to the judges' assessments of the final revised version of the tool, the kappa values rose to 0.80, indicating substantial agreement.

[Table 3 near here]

Given these inter-agreement results, we concluded that the dimensions, variables and indicators of the revised version of the tool, which are presented in Table 1, were appropriate for our purposes.

3.2 Inter-observer reliability

The results of the inter-observer reliability analysis using Krippendorff's alpha (Krippendorff, 2004) are shown in Table 4. As it exceeded the $\alpha \geq .800$ criterion level, we concluded that no further training was needed.

[Table 4 near here]

3.3 Descriptive results for activity and inactivity

A summary of the descriptive results from the 18 recorded segments analysed in terms of the dimensions of 'activity' (variables: listening, oral interaction and other activity) and 'inactivity' (variables: motor locomotion and motor locomotion + L2 language skills) appears in Table 5, which shows minimum and maximum values expressed in seconds, percentages of time, means and standard deviations. The results indicate that, on average, 41.86% of class time in the PE-in-CLIL lessons in this corpus was taken up with motor activities (dimension: activity time). Of this motor-engaged time, 46.01% corresponded to the variable 'motor locomotion' and 53.98% corresponded to the variable 'motor locomotion + L2 language skills', meaning that during this time students were communicating in the L2 with the teacher or their peers even though it was not a part of the task requirements and they were doing so without hindering the pace of the game or physical activity. By contrast, the total amount of inactivity time (dimension: inactivity) was 58.12%, of which 33.20% corresponded to 'listening', 44.56% to 'oral interaction' and 22.22% to 'other types of inactivity'.

[Table 5 near here]

3.4 Descriptive results for lesson context variables

Table 6 illustrates the lesson context variables. Under 'educational level', ten of the 18 recordings (55.6%) were recorded in a primary education context and the remaining eight (44.4%) in a secondary education context. The 'type of physical activity' was distributed as follows: games were present in eight videos, body language, bodily expression, dance and rhythmic activities in four and fundamental motor skills in six. With respect to the use of language support materials to facilitate communication, this was present in 13 of the 18 recordings (72.2%). When it comes to language skills,

speaking activities were performed while doing a task in four videos, reading activities before doing a task were observed in ten and writing activities were present in only four.

[Table 6 near here]

3.5 Relationships between lesson context and activity/inactivity dimensions

When subjected to a Mann-Whitney U test, variables such as educational level ($U = -.800$, $p = .424$), speaking ($H(2) = 0.821$, $p = .663$) and writing ($U = -.425$, $p = .671$) did not show statistically meaningful differences across the two ‘Activity’/‘Inactivity’ dimensions [?], meaning that the levels of inactivity and activity in motor tasks did not vary significantly with those contextual variables. However, the relationship between activity/inactivity and the type of physical activity, the use of language support materials, and reading did reveal significant results, which are detailed below.

3.5.1 Use of language support materials

The Mann-Whitney U test indicated that the use of language support materials such as flashcards to facilitate communication lowered activity time (33.11%) and increased inactivity time (66.88%) ($U = -2.809$, $p = .005$). Therefore motor-engaged time was affected by whether such materials were used during PE lessons (table 7).

[Table 7 near here]

3.5.2 Types of physical activity

There was a statistically significant relationship between the types of physical activity and activity time ($H(2) = 6.703$, $P = .035$), with a mean of 61.19% of activity time when the type of physical activity was ‘body language, bodily expression, dance and rhythmic activities’, 42.01% of activity time when the type was ‘games’ and 28.79% when it was ‘fundamental motor skills’ (table 8). Therefore activity time was most affected when fundamental motor skills were being taught.

[Table 8 near here]

3.5.3 Reading

A Kruskal-Wallis test showed that there was a statistically significant relationship between reading and activity time ($H(2) = 7.232, p = .027$), with a mean of 52.90% of activity time when no reading was present in the task, 32.69% of activity time when reading was present either before, during or after the task and 20.37% of activity time when reading was present in at least two segments of the task (i.e. both before and during) (Table 9). This suggests that, of the three skills, reading most significantly reduced activity time.

[Table 9 near here]

4. Discussion

This study was inspired by the need to provide PE-in-CLIL teachers and researchers with a validated tool to observe PE lessons taught in a foreign language. Our results point to the successful accomplishment of this goal. To answer RQ1, our decision to base the new tool on the dimensions ‘activity’ and ‘inactivity’ are supported by the fact that these two dimensions have been used successfully in previous studies to observe and analyse PE lessons (see, among others, Siedentop et al. 1982; McKenzie et al. 1991). However, unlike in previous studies, in our study the variables comprising these two dimensions had to include not only physical activity but also L2 language use. Thus, we simplified the behavioural classification presented by Roberts and Fairclough (2012) and then combined it with language aspects that are present in classroom interactional competence, sheltered instruction (Echevarría et al. 2010) and CLIL settings (Coyle et al. 2010), which left us with five variables (listening, oral interaction, other inactivity, motor locomotion only and motor locomotion + language skills) along with their corresponding indicators. Furthermore, following the example provided by Roberts and Fairclough (2012), a number of contextual variables concerning both PE and language were included in a supplementary checklist to gather information to

correlate with the abovementioned variables and therefore provide clues about possible relationships among them.

In answer to RQ2, previous research informed our choice of dimensions, variables and indicators and likewise encouraged us to carry out a validity rating process using a four-point Likert scale. The process of observer training was based on the work of Losada and Manolov (2015) and the choice of the LINCE software was informed by Gabin et al. (2012). In this case, a two-step procedure consisting of basic training and applied training was consistent with the training of experienced PE-in-CLIL teachers who, although they had mastered the use of information and communication tools in education, were unfamiliar with the observation instrument. With regard to RQ3, content validity was confirmed, with the final tests yielding an inter-judge agreement percentage of 93.33 and kappa values of 0.80. The reliability of the observer training was also confirmed as Krippendorff's alpha (ratio) yielded a value of 0.932, indicating that no further training was needed.

Concerning RQ4, the results suggest that pupils spend less time engaged in physical activity when PE is taught in a L2 than what is seen in similar studies involving PE taught in a L1. Our pilot test showed students spending 41.86% of their time doing motor activities, which contrasts with the 50% of time that children should be physically active during a PE lesson according to the recommendations of the US Department of Health and Human Services (Bevans et al. 2010) or the typical PE average (49.7%) reported by Logan et al. (2015). It is also much lower than the duration MVPA during mastery (68%) and performance (67%) climates reported by Logan et al. (2015) and Martínez-Hita and García-Canto (2017), who observed physical activity during 65.35% of lesson time when PE was taught in L1. When the relationship between activity time and contextual factors was analysed, fundamental motor skills seemed to be most negatively impacted by CLIL activities because the average amount of time in which motor skill practice could be observed was only 28.79%. In addition, the use of support materials like flashcards to facilitate communication in L2 significantly reduced children's activity time as well (33.11%). Finally, when any kind of reading activity occurred before, during or after the lesson, the percentage of the total activity time was between 20.37% and 32.69%.

Although such findings cannot be compared with previous studies due to the lack of previous research on PE-in-CLIL, they should clearly be a cause for some concern. In a general sense, if physical activity is used to make language learning more interesting and motivating (Zindler 2013; Lynott 2008; Coral 2010) by embedding it in a PE context and yet this approach jeopardises PE contents, this suggests that the PE classes are being ‘instrumentalised’ for the purpose of language learning, thus losing part of their primary function. However, according to Coyle et al. (2010), this is not what underpins the principles of the CLIL approach, where a balance between content and language must be achieved (Coral and Lleixà 2016).

One important implication of these findings is that the amount of school time allotted to PE must be given careful consideration when schools set up a PE-in-CLIL programme. Earlier PE-in-CLIL programmes (Coral 2010) proactively took measures to increase the time allotted to PE by 25% and five years later the programme showed outstanding L2 results (Coral et al. 2016) without jeopardising PE contents. Another option is to gain extra time for the subject by including PE in school CLIL projects (Zindler 2013). A second important implication of these results is the need for specific PE-in-CLIL teacher training. In this regard, Coral and Lleixà (2017) point out that more attention should be paid to methodological aspects in training to help new PE-in-CLIL teachers know how to (a) balance content and language, (b) embed the L2 in physical activity without slowing the pace of the task and (c) apply language and content scaffolding strategies in association with PE teaching styles to bridge learning gaps.

5. Conclusions

We have described how, following a four-step procedure, a PE-in-CLIL observation tool was first validated for content validity and internal consistency and then successfully tested, thus demonstrating the tool’s efficacy for recording child activity and analysing lesson context while taking language aspects into account. The pilot testing was carried out using recordings of real classroom lessons being taught in natural contexts by PE-in-CLIL teachers at schools, thus lending the findings ecological validity. At the same time, the results of the pilot testing revealed that in the PE-in-CLIL classes analysed the children’s physical activity time was slightly lower than what is recommended for PE classes by health authorities and previous studies. This shortfall

suggests that it is difficult for PE-in-CLIL teachers to maintain an appropriate balance between promoting physical activity and fostering L2 language skills. For example, the use of language support materials like flashcards may well facilitate L2 learning, but at the same time it takes times which might otherwise be devoted to motor activity. The same may be said for any kind of reading activity in the PE context. Two strategies to help address this issue have been proposed, namely better training for PE-in-CLIL teachers and the allocation of longer time periods to PE-in-CLIL classes relative to non-CLIL PE classes. However, given the small sample size used in the pilot study, caution must be advised about drawing broad conclusions on the basis of the results reported here. A good deal more research on this subject needs to be undertaken before the association between physical activity time and L2 learning in PE-in-CLIL classes can be fully understood. However, the observation tool described here will make such research much easier to carry out.

6. Acknowledgments

The authors acknowledge the collaboration of the teachers and schools who participated in the study and the support provided by the *Centre de Recursos Pedagògics Específics de Suport a la Innovació i la Recerca Educativa* (Centre for Teaching Resources for Innovation and Research in Education, CESIRE) of the Ministry of Education of the Government of Catalonia, by the research group *Llengua i Educació* (Language and Education, LED) (ref. 2014S GR1190) of the *Universitat Autònoma de Barcelona*, and by the research group *l'Educació Física i l'Esport* (Physical Education and Sport, EDUFISES) of the *Universitat de Barcelona*.

7. References

- Admiral, Wilfried, Gerard Westhoff, and Kees de Bot. 2006. "Evaluation of Bilingual Secondary Education in The Netherlands: Students' language proficiency in English." *Education Research and Evaluation* 12 (1): 75-93. doi: 10.1080/13803610500392160.
- Altman, D.G. 1991. *Practical Statistics for Medical Research*. London England: Chapman and Hall.

- Australian Curriculum, Assessment and Reporting Authority. 2012. *The shape of the Australian Curriculum: Health and Physical Education*. Available at: <https://www.acara.edu.au/> (accessed 10 January 2017).
- Coral, Josep. 2010. “L’aprenentatge de l’anglès a través de l’educació física: el programa ‘Mou-te i aprèn’”. *Temps d’Educació*, no. 39: 149–170.
- Coral, Josep. 2013. “Physical education and English integrated learning: How school teachers can develop PE-in-CLIL programmes”. *Temps d’Educació*, no. 45: 41–65.
- Coral, Josep and Teresa Lleixà. 2013. “Las tareas en el aprendizaje integrado de educación física y lengua extranjera (AICLE). Determinación de las características de las tareas mediante el análisis del diario de clase”. *Retos. Nuevas Tendencias en Educación Física, Deporte y Recreación*, no. 24: 79–84.
- Coral, Josep and Teresa Lleixà. 2016. “Physical education in content and language integrated learning: successful interaction between physical education and English as a foreign language”. *International Journal of Bilingual Education and Bilingualism* 19(1):108-126. doi:10.1080/13670050.2014.977766.
- Coral, Josep and Teresa Lleixà. 2017. In-service Content and Language Integrated Learning (CLIL) teachers: An Action Research Project in professional learning. In: *Teachers and Teacher Educators Learning Through Inquiry: International Perspectives* edited by P. Boyd and A. Szplit, 255-279. Libron: Kraków.
- Coral, Josep, Teresa Lleixà and Carles Ventura. 2016. “Foreign language competence and content and language integrated learning in multilingual schools in Catalonia: an ex-pot-facto study analysing the results of state key competences testing”. *International Journal of Bilingual Education and Bilingualism*. doi:10.1080/13670050.2016.1143445.
- Lleixà, Teresa, Marta Capllonch, and Carles Gonzàlez. 2015. “Competencias básicas y programación de Educación Física. Validación de un cuestionario diagnóstico”. *Retos. Nuevas Tendencias en educación Física, Deporte y Recreación*, no. 27: 52-57.
- Bell, Nancy D., and David Lorenzi. 2004. “Facilitating Second Language Acquisition in Elementary and Secondary Physical Education Classes”. *Journal of Physical Education, Recreation & Dance* 75 (6): 46-51. doi:10.1080/07303084.2004.10607257.

- Behets, Daniel. 1997. "Comparison of more and less effective teaching behaviors in Secondary physical education". *Teaching and Teacher Education* 13 (2): 215-224. doi: 10.1016/S0742-051X(96)00015-7.
- Bevans, Katherine, L.A Fitzpatrick, Betty Sánchez, and Christopher B. Forrest. 2010. "Individual and Instructional Determinants of Student Engagement in Physical Education". *Journal of Teaching in Physical Education* 29 (4): 399-416.
- Bonnet, Andreas. 2012. "Towards an Evidence Base for CLIL. How to Integrate Qualitative and Quantitative as well as Process, Product and Participant Perspectives in CLIL Research". *International CLIL Research Journal* 1 (4): 66-78.
- Bruton, Anthony. 2011. "Is CLIL so beneficial, or just selective? Re-evaluating some of the research." *System* 39 (4): 523–532. doi: 10.1016/j.system.2011.08.002.
- Chow, Bik. C., Thomas L. McKenzie, and Lobo Louie. 2015. "Children's Physical Activity and Associated Variables during Preschool Physical Education". *Advances in Physical Education* 5: 39-49. doi: 10.4236/ape.2015.51005.
- Coyle, Do, Philip Hood, and David Marsh. 2010. *Content and Language Integrated Learning*. Cambridge: Cambridge University Press.
- Dalton-Puffer, Christiane. 2011. "Content and Language Integrated Learning: From Practice to Principles?" *Annual Review of Applied Linguistics* 31: 182–204. doi: 10.1017/S02671905110000092.
- Derri, Vassiliki, Kyriaki Emmanouilidou, Olga Vassiliadou, Efthimis Kioumourtzoglou, and Edmundo Loza. 2007. "Academic learning time in physical education (ALT-PE): is it related to fundamental movement skill acquisition and learning?" *Ricyde* 6 (3): 12-23. doi: 10.5232/ricyde2007.00602.
- Derri, Vassiliki, Kyriaki Emmanouilidou, Olga Vassiliadou, George Tzetzis, and Efthimis Kioumourtzoglou. 2008. "Relationship between academic learning time in physical education (ALT-PE) and skill concepts acquisition and retention". *Physical Educator* 65 (3): 134-145.
- Devos, Nathan J. 2012. Content and Language Integrated Learning in physical education: evidence for language and content scaffolding during peer interaction. In: *Learner Autonomy in the English Classroom. Empirical Studies and Ideas for Teachers* edited by P. Lennon, 359-381. Frankfurt: Peter Lang.
- Devos, Nathan J. 2016. *Peer Interactions in New Content and Language Integrated Settings*. Heidelberg: Springer.

- Echevarría, Jana J., M.E. Vogt, and D.J. Short. 2010. *Making Content Comprehensible for English Learners*. Boston, MA: Pearson.
- Erfle, Stephen, and Abigail Gamble. 2014. "Effects of Daily Physical Education on Physical Fitness and Weight Status in Middle School Adolescents". *Journal of School Health* 85 (1): 27-35.
- Fernández-Sanjurjo, Javier, Alberto Fernández-Costales, and J.M. Arias-Blanco. 2017. "Analysing students' content-learning in science in CLIL vs. non-CLIL programmes: empirical evidence from Spain". *International Journal of Bilingual Education and Bilingualism*. doi: 10.1080/13670050.2017.1294142.
- Gajo, Laurent. 2007. "Linguistic knowledge and Subject knowledge: how does bilingualism contribute to subject development?" *International Journal of Bilingual Education and Bilingualism* 10 (5): 563-581.
- Gabin, Brais, Oleguer Camerino, M.T. Anguera, and Marta Castanyer. 2012. "Lince: multiplatform sport analysis software". *Procedia-Social and Behavioral Sciences* no. 46: 4692-4694. doi: 10.1016/j.sbspro.2012.06.320.
- Generalitat de Catalunya (2015) DECRET 119/2015, de 23 de juny, d'ordenació dels ensenyaments de l'educació primària. [Ordinance of primary education teaching in Catalonia] (DOG Núm. 6900 - 26.6.2015)
- Generalitat de Catalunya. (2016) *Competències bàsiques de l'àmbit de l'educació física*. Available at <http://ensenyament.gencat.cat/web/.content/home/departament/publicacions/collections/competencies-basiques/primaria/prim-educacio-fisica.pdf> (accessed 11 January 2017).
- Harvey, Stephen, Yang Song, Junhung Baek, and Hans van der Mars. 2016. "Two sides of the same coin: Student physical activity levels during a game-centred soccer unit". *European Physical Education Review* 22 (4): 411-429. doi:10.1177/1356336X15614783.
- Heidorn, Brent, R. Glenn Weaver, and Aaron Beighle, A. 2016. "Physical Activity and Physical Education: A Combined Approach". *Journal of Physical Education, Recreation & Dance* 7 (4): 6-7. doi: 10.1080/07303084.2016.1142184.
- Hollis J.L., A.J. Williams, R. Sutherland, E. Campbell, N. Nathan, L. Wolfenden, P.J. Morgan et al. 2015. "A systematic review and meta-analysis of moderate-to-vigorous physical activity levels in elementary school physical education lessons". *Preventive Medicine* 86: 34-54. doi: 10.1016/j.ypmed.2015.11.018.

- Harro, Maarike. 1997. "Validation of a Questionnaire to Assess Physical Activity of Children Ages 4-8 Years". *Research Quarterly for Exercise and Sport* 68(4): 259-268. doi:10.1080/02701367.1997.10608007
- House, A.E., B.J. Houseand, and M.B. Campbell .1981. "Measures of interobserver agreement: Calculation formulas and distribution effects". *Journal of Behavioral Assessment* 3(1): 37-57.
- Ioannou-Georgiou, Sophie. 2012. "Reviewing the puzzle of CLIL". *ELT Journal* 66 (4): 495–504. doi:10.1093/elt/ccs047.
- Janssen, Ian, and A.G. LeBlanc. 2010. "Systematic review of the health benefits of physical activity and fitness in school-aged children and youth". *International Journal of Behavioral Nutrition and Physical Activity* 7 (40): 1-16. doi:10.1186/1479-5868-7-40.
- Krippendorff, Klaus. 2004. *Content Analysis: An Introduction to Its Methodology* (2nd ed.). Thousand Oaks, CA: Sage.
- Landis, J. Richard, and G.G. Koch. 1977. "The measurement of observer agreement for categorical data". *Biometrics* 33(1): 159–174. doi: 10.2307/2529310
- Lasagabaster, David. 2008. "Foreign Language Competence in Content and Language Integrated Courses". *The Open Applied Linguistics Journal*, no.1: 31–42. doi:10.2174/1874913500801010030.
- Logan, Samuel W., L.E. Robinson, E.K. Webster, and M.E. Rudisill. 2015 "The influence of instructional climates on time spent in management tasks and physical activity of 2nd-grade students during physical education". *European Physical Education Review*. 21 (2): 195–205. doi: /doi.org/10.1177/1356336X14555304.
- Losada, J.Luis, and Rumen Manolov. 2015. "The process of basic training, applied training, maintaining the performance of an observer". *Qual Quant* 49:339-347. doi: /doi.org/10.1007/s11135-014-9989-7.
- Lorenzo, Francisco, Sonia Casal, and Pat Moore. 2009. "The Effects of Content and Language Integrated Learning in European Education: Key Findings from Andalusian Bilingual Sections Evaluation Project". *Applied Linguistics* 31 (3): 418–442. doi: 10.1093/applin/amp041.
- Lozano, Luis, Jesús Viciano, and Maurice Piéron. M. 2006. "Anàlisi dels instruments d'observació utilitzats per al registre de variables temporals en educació física". *Revista Apunts*, no 84: 22-31.

- Lynott, Francis J. 2008. “Integrating Other Subject Matter Without Jeopardizing Physical Education Goals: The Content Linkage Approach”. *Strategies* 22 (1): 10-17. doi:10.1080/08924562.2008.10590802.
- Martínez-Hita, Francisco J., and Eliseo García-Canto. 2017. “Influencia del bilingüismo en el tiempo de compromiso motor en Educación Física”. *Retos*, no. 32: 178-182.
- McKenzie, T.L., J.F. Sallis, and P.R. Nader. 1991. “SOFIT: System for observing fitness instruction time”. *Journal of Teaching in Physical Education* 11 (2): 195-205. doi:10.1123/jtpe.11.2.195.
- Metzler, Michael. 1989. “A review of research on time in sport pedagogy”. *Journal of Teaching in Physical Education* 8 (2): 87-103. doi: 10.1123/jtpe.8.2.87.
- Meyer, Oliver. 2010. “Towards Quality-CLIL: Successful Planning and Teaching Strategies”. *Pulso*, no. 33:11–29.
- Navés, Teresa. 2009. Effective Content and language Integrated Learning (CLIL) Programmes. In: *Content and Language Integrated Learning. Evidence from Research in Europe* edited by Yolanda Ruiz de Zarobe and R.M. Jiménez-Catalán, 22-40. Clevedon: Multilingual Matters.
- Olmedo, J.A. 2000. “Estrategias para aumentar el tiempo de práctica motriz en las clases de Educación Física escolar”. *Apunts*, no. 59: 22-30.
- Quivy, Raymond, and Luc Van Carnpenhoudt. 2005. *Manual de investigación en ciencias sociales*. México: Limusa.
- Paran, Amos. 2013. Content and Language Integrated Learning: Panacea or Policy Borrowing Myth? *Applied Linguistics Review* 4(2): 317–342. doi:10.1515/applirev-2013-0014
- Pate RR, M.G. Davis, T.N. Robinson, E.J. Stone, T.L. McKenzie, and J.C. Young. 2006. “Promoting physical activity in children and youth: a leadership role for schools: a scientific statement from the American Heart Association Council on Nutrition, Physical Activity, and Metabolism (Physical Activity Committee) in collaboration with the Councils on Cardiovascular Disease in the Young and Cardiovascular Nursing”. *Circulation*, no. 114: 1214–1224.
- Pérez-Cañado, Maria Luisa. 2012. “CLIL research in Europe: past, present and future”. *International Journal of Bilingual Education and Bilingualism* 15 (3): 315-341. doi:10.1080/13670050.2011.630064.

- Pieron, Maurice. (1999). *Por una enseñanza eficaz de la educación física*. Barcelona, INDE.
- Roberts, Simon, and Stuart Fairclough. 2012. "A five-stage process for the development and validation of a systematic observation instrument: The system for observing the teaching of games in physical education (SOTG-PE)". *European Physical Education Review* 18(1): 97-113. doi: 10.1177/1356336X11430653.
- Rottmann, Birte. 2006. *Sport auf Englisch: Lerngelegenheiten im bilingualen Sportunterricht*, PhD Thesis, University of Hamburg.
- Senne, Terry, David Rowe, Boni Boswell, James Decker, and Shaun Douglas. 2010. "Factors associated with adolescent physical activity during middle school physical education: A one-year case study". *European Physical Education Review*, 15 (3): 295–314. doi: 10.1177/1356336X09364722.
- Scottish Government. 2004. *Curriculum for excellence: health and wellbeing. Experiences and outcomes*. Available at: <https://www.education.gov.scot/Documents/health-and-wellbeing-eo.pdf> (accessed 10 January 2017)
- Siedentop, Daryl, Marielle Tousignant, and Melissa Parker 1982. *Academic Learning Time-Physical Education Coding Manual*. Columbus OH: Ohio State University.
- Silverman, Stephen. 1985. "Student characteristics mediating engagement – outcome relationships in PE". *Research Quarterly for Exercise and Sport* 56(1): 66-72. doi: 10.1080/02701367.1985.10608433.
- Simón, J.A., J.G. Fernández, and O.R. Contreras, 2017. "Diseño y validación de un cuestionario de autopercepción de la excelencia en el deporte". *Retos* 31: 58-63.
- Smith, N.J., S.M. Monnat, and M. Lounsbury, 2015. "Physical activity in physical education: are longer lessons better?" *Journal of School Health* 85:141-148. doi: 10.1111/josh.12233.
- Surmont, Jill, Esli Struys, Van Den Noort, Maurits and Piet Van De Craen. 2016. The effects of CLIL on mathematical content learning: A longitudinal study. *Studies in Second Language Learning and Teaching*, 6(2), 319-337. doi:10.14746/ssllt.2016.6.2.7.
- Sylvén, L.K. 2013. "CLIL in Sweden – why does it not work? A metaperspective on CLIL across contexts in Europe." *International Journal of Bilingual Education and Bilingualism* 16 (3): 301–320. doi:10.180/13670050.2013.777387.

- Van de Craen, P, E. Ceuleers & K. Mondt. 2007. 'Cognitive development and bilingualism in primary schools: teaching maths in a CLIL environment'. In: D. Marsh & D. Wolff (eds) *Diverse Contexts Convergging Goals. CLIL in Europe*. Frankfurt am Main: Peter Lang, 185-200.
- Viciano, Jesus, Luis Lozano, Armando Cocca, and Daniel Mayorga. 2012. "Influence of the Organizational System on Motor Engagement Time in Physical Education on High School Students". *Procedia. Social and Behavioral Sciences*, 69:1160-1167. doi: 10.1016/j.sbspro.2012.12.046.
- Weaver, R.G. 2016. "SOFIT+: An Expanded System to Capture PE-Based Physical Activity Promotion". *Research Quarterly for Exercise and Sport*, no. 87, Supplement 2, A4-A5.
- World Health Organization. 2010. *Global recommendations on physical activity for health*. Geneva: WHO Press. Available at: http://apps.who.int/iris/bitstream/10665/44399/1/9789241599979_eng.pdf (accessed 10 January 2017)
- Yanci, Javier, Anabel Vinuesa, Josune Rodriguez, Leire Yanci. 2016. "El tiempo de compromiso motor en las sesiones de Educación Física del primer y segundo ciclo de Educación Primaria". *Sportis. Scientific Journal of School Sport, Physical Education and Psychomotricity* 2(2): 239-253. doi: 10.17979/sportis.2016.2.2.1447.
- Zindler, Kerstin. 2013. *Content and Language Integrated Learning (CLIL) and PE in England. An explanatory Study*, PhD Thesis, University of Sheffield.

Table 1. Dimensions, variables and indicators in the observation tool used to analyse PE-in-CLIL classes. The codes assigned to each variable were used to label excerpts from video recordings

Dimension	Variables	Code	Indicators
	Listening	LI	<p>Class time where children are inactive while listening to the teacher, e.g.:</p> <p>a) teacher talking and students just listening (not asking or answering).</p> <p>b) teacher talking, and when a child claims not to understand the explanation, the teacher changes or adjusts the explanation without achieving a minimum level of oral interaction.</p>
Inactivity	Oral interaction	OI	<p>Class time where children are inactive while any type of oral interaction related to the topic is occurring, e.g.:</p> <p>a) teacher explaining and students asking.</p> <p>b) students asking and teacher answering.</p> <p>c) teacher asking students to show comprehension.</p> <p>d) students asking for clarification and teacher reformulating the explanation.</p> <p>e) students asking for examples and teacher or other students modelling through a physical answer to support the oral explanation.</p>
	Other	Other	<p>Other class time where children are physically inactive (forming a queue, putting on vests, ending the explanation and starting the activity, pauses during the explanation, etc.) or engaged in an off-task, non-learning motor or language activity (e.g. not part of the planned lesson activity, such as waiting to start the lesson or chasing somebody before starting the lesson).</p>

Activity	Motor-locomotion only	MR-LO	Class time where children are: a) performing a game-related motor response (e.g. throwing or catching). b) engaged in on-task performance of a locomotor task that was part of a practice, drill or game (e.g. running or jumping). c) engaged in on-task performance of a locomotive task while performing a game-related motor response (e.g. jumping while shooting or running while dribbling). d) performing any other type of physical activity (e.g. stretching or dancing).
	Motor-locomotion + L2 language skills	MR-LO + LANG	Class time where children are engaged on-task in motor-response, locomotor activity or any other type of physical activity where saying a password, pass-phrase, pass-sentence, giving feedback or any other oral act that is part of the task is required and performed without stopping the game or physical activity.

Table 2. Supplementary checklist of variables related to the context in which the observed lesson was taking place.

Variables of lesson context	Descriptors
Educational level	Primary
	Secondary
Type of physical activities	Games: Minor games (games that involve simple and variable rules such as tag games, traditional games, hit-the-target games, etc.), sports-oriented games and modified games.
	Body language, bodily expression, dance and rhythmic activities.
	Fundamental motor skills: Locomotor, stability and manipulation activities.
Use of language support materials to facilitate communication	Yes. Children are provided with flashcards or other written language support materials to facilitate communication.
	No. Children are not provided with flashcards or other written language support materials to facilitate communication.
Speaking (0 out of 3; 1 out of 3; 2/3 out of 3)	Before: Speaking activities are included before doing a motor task.
	During: Speaking activities are done while performing a motor task.
	After: Speaking activities are included after doing a motor task.
Reading (0 out of 3; 1 out of 3; 2/3 out of 3)	Before: Reading activities are included before doing a motor task
	During: Reading activities are done while performing a motor task.

	After: Reading activities are included after doing a motor task.
Writing (yes or no)	Writing activities are included in the lesson.

Table 3. Percentages of agreement between judges' assessments of the preliminary and final versions of the observation tool according to the formula proposed by House et al. (1981) and kappa values.

Dimensions	Variables	First version	Second version
Inactivity	Listening	66.66	100
	Oral interaction	66.66	83.33
	Other	83.33	100
Activity	Motor Locomotion only	83.33	100
	Motor Locomotion + Language Skills	83.33	83.33
Average		76.66	93.33
Kappa		0.56	0.80

Table 4. Inter-observer reliability using Krippendorff's alpha between the three observers following basic and applied training.

N coders	N cases	N decisions	Krippendorff's alpha (ratio)
3	10	30	0.932

Table 5. Descriptive statistics for the dimensions 'activity' and 'inactivity' as observed in 18 recorded PE-in-CLIL class sessions. For each item, top row values are expressed in seconds and bottom row values refer to the percentage of class time spent on this variable.

Descriptive statistics, activity and inactivity dimensions					
	N	Minimum	Maximum	M	SD
Listening	18	63.70	763.00	338.26	195.27
% Listening	18	3.77	71.19	33.20	15.51
Oral Interaction	18	71.70	1345.00	526.80	382.43
% Oral Interaction	18	12.15	73.07	44.56	16.39
Other Inactivity	18	20.00	1317.20	280.58	316.41
% Other Inactivity	18	3.13	55.11	22.22	15.24
Total Inactivity	18	330.00	2479.00	1145.66	658.70
% Total Inactivity	18	28.88	95.37	58.12	20.10
Motor Locomotion	18	.00	3112.00	555.61	835.20
% Motor Locomotion	18	.00	100.00	46.01	32.89

Motor Locomotion + L2 Language Skills	18	.00	978.00	342.79	293.87
% Motor Locomotion + L2 Language Skills	18	.00	100.00	53.98	32.89
Total Activity	18	116.00	3145.00	898.38	813.97
% Total Activity	18	4.63	71.12	41.86	20.10

Table 6. Descriptive statistics related to lesson context variables of the 18 video-recorded PE-in-CLIL lessons used in the pilot testing.

Lesson context variables	Descriptors	Number of recordings	Percentage (out of 18 recordings)
Educational level	Primary	10	55.6
	Secondary	8	44.4
Type of physical activities	Games: Minor games (games that involve simple and variable rules such as tag games, traditional games, target games, etc.), sports-oriented games and modified games.	8	44.4
	Body language, bodily expression, dance and rhythmic activities.	4	22.2
	Fundamental motor skills: Locomotor, stability and manipulation activities.	6	33.3
Use of language support materials to facilitate communication	Yes. Children are provided with flashcards or other written language support materials to facilitate communication.	13	72.2
	No. Children are not provided with flashcards or other written language support materials to facilitate communication.	5	27.8
Speaking	Before: Speaking activities are included before doing a motor task.	5	27.8
	During: Speaking activities are done while performing a motor task.	8	44.4
	After: Speaking activities are included after doing a motor task.	5	27.8
Reading	Before: Reading activities are included before doing a motor task.	10	55.6
	During: Reading activities are done while performing a task.	5	27.8
	After: Reading activities are included after doing a task.	3	16.7
Writing	Included	4	22.2
	Not included	14	77.8

Table 7. Mean percentage of inactivity and activity time relative to the use of language support materials to facilitate communication.

	Language support materials used	
	Yes	No
% Total Inactivity time	66.88	35.38
% Total Activity time	33.11	64.62

Table 8. Percentage of inactivity and activity time considering the type of physical activity.

	Type of physical activity		
	Games	Body language	Fundamental motor skills
% Total Inactivity time	57.98	38.81	71.21
% Total Activity time	42.01	61.19	28.79

Table 9. Percentage of total inactivity and activity time during which activities were present. '0 out of 3' means that no reading activities were present during the task. '1 out of 3' means that reading activities took place either before, during or after the task. '2/3 out of 3' means that reading took places in any two phases of the task or throughout it.

	Reading: before, during or after the task.		
	0 out of 3	1 out of 3	2/3 out of 3
% Total Inactivity	47.10	67.29	79.63
% Total Activity	52.90	32.69	20.37