UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Practical solution or missed opportunity? The impact of language of instruction on history teachers' application of pedagogical content knowledge (PCK)

Oattes, H.; Wilschut, A.; Oostdam, R.; Fukkink, R.; de Graaff, R.

DOI
10.1016/j.tate.2022.103721

Publication date
2022

Document Version
Final published version

Published in
Teaching and Teacher Education

License
CC BY

Link to publication

Citation for published version (APA):
Oattes, H., Wilschut, A., Oostdam, R., Fukkink, R., & de Graaff, R. (2022). Practical solution or missed opportunity? The impact of language of instruction on history teachers' application of pedagogical content knowledge (PCK). *Teaching and Teacher Education*, *115*, Article 103721. https://doi.org/10.1016/j.tate.2022.103721

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:17 Nov 2026

Teaching and Teacher Education 115 (2022) 103721

Contents lists available at ScienceDirect

Teaching and Teacher Education

journal homepage: www.elsevier.com/locate/tate

Research paper

Practical solution or missed opportunity? The impact of language of instruction on Dutch history teachers' application of pedagogical content knowledge (PCK)

Huub Oattes [a, *], Arie Wilschut [a], Ron Oostdam [b], Ruben Fukkink [a], Rick de Graaff [c]

[a] Amsterdam University of Applied Sciences, Faculty of Education, Centre for Applied Research in Education (CARE), Wibautstraat 2-4, 1091, GM, Amsterdam, the Netherlands
[b] Amsterdam University of Applied Sciences, University of Amsterdam, Research Institute of Child Development and Education, Nieuwe Achtergracht 127, 1018, WS, Amsterdam, the Netherlands
[c] University of Utrecht, Faculty of Humanities, Languages and Communication, Trans 10, 3512, JK, Utrecht, the Netherlands

HIGHLIGHTS

- Teachers' use of PCK shows strong similarity irrespective of the languages used.
- Teachers' use of PCK is more advanced in grades 9.
- Teachers' PCK in mainstream and bilingual education can be improved.

ARTICLE INFO

Article history:
Received 30 September 2020
Received in revised form
15 February 2022
Accepted 22 March 2022
Available online xxx

Keywords:
Junior secondary education
History teacher
Pedagogical content knowledge
Bilingual history education
CLIL

ABSTRACT

Teaching history requires clear, detailed and subject specific language. History teachers teaching in a second language are confronted with students' second language limitations, which likely have an aggravating impact on their application of pedagogical content knowledge (PCK). We analysed and compared 12 Dutch spoken and 12 English spoken paired history lessons in junior grades 7 and 9. Contrary to our expectation, we found a strong similarity of the teachers' PCK application in both grades 7 and 9, irrespective of the used language. The PCK application in both grades and languages was of average quality, while the PCK used in grade 9 was more advanced.

© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

During the last 25 years, globalisation and the digital revolution have had an enormous impact on the life of many people. The need for a lingua franca to participate in this global community has stimulated many countries to include English as a foreign language in their countries' language policy and school curricula. As a result of the Barcelona European Council meeting in 2002, the European Union encouraged its member states to teach at least two foreign languages from a very early age (European Commission, 2009). In the EU countries, English (97%) is the most learned language (French 34%, German 23%) in junior secondary education (European Commission/EACEA/Eurydice, 2017). A growing number of European schools apply forms of Bilingual Education (BE) to intensify the process of foreign language learning. They are encouraged by research showing that the development of proficiency in multiple languages is possible, which makes it attractive for policy makers, educators, and parents (Tucker, 2001).

In bilingual education several teaching approaches can be distinguished. While in secondary education in European countries BE is mainly practiced through Content and Language Integrated Learning (CLIL), in other contexts Content-Based Instruction (CBI) or English as a Medium of Instruction (EMI) are popular bilingual

* Corresponding author.
E-mail addresses: h.oattes@hva.nl (H. Oattes), a.h.j.wilschut@hva.nl (A. Wilschut), r.j.oostdam@uva.nl (R. Oostdam), r.g.fukkink@uva.nl (R. Fukkink), r.degraaff@uu.nl (R. de Graaff).

https://doi.org/10.1016/j.tate.2022.103721
0742-051X/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

approaches (Lin, 2016). CLIL, CBI and EMI lack a crystal-clear definition, but share the basic concept of the use of a second or additional language as the medium of instruction of academic content. Cenoz (2015) insists there are no substantial differences between CLIL and CBI, the context, and the typical type of CLIL/CBI/ student. We, on the other hand, believe that a major difference between the two can be found in the background and context of participating students and teachers. In CBI teachers are usually native speakers of English who teach immigrant students the language of the majority out of necessity to assist with their successful integration in society (Jaffee, 2016; Sclafani, 2017). In EMI English is usually chosen as a medium of instruction to attract international students and teachers to a program, no matter if they are speakers of English as first (L1) or second language (L2).

Dutch CLIL teachers are usually non-native English speakers, and, like the students, members of the majority population group. CLIL teachers and students voluntarily choose to participate in BE because they are looking for an educational challenge to improve their second language and future study and/or work prospects. CLIL appears in numerous forms and shapes depending on a country's foreign language learning policy (Dalton-Puffer & Nikula, 2014).[1] In this study we define CLIL as a dual focused educational approach in which an additional language is used for the learning and teaching of both content and language. That is, in the teaching and learning process there is a focus not only on content and not only on language. Each is interwoven, even if the emphasis is greater on one or the other at a given time (Coyle, Hood, & Marsh, 2010, p. 1).

Research into the effects of BE has primarily concentrated on the students and their development of the second language proficiency, and much less on their subject content knowledge progression. The effects on subject teachers of teaching content and language in two languages has hardly been researched (Cammarata & Tedick, 2012; Coonan, 2007; Moate, 2011; Oattes et al., 2018a, 2018b; Papaja, 2013), despite their pivotal role in CLIL (Dalton-Puffer & Smit, 2013; Pérez-Cañado, 2012). We want to establish how CLIL teachers teach history content in English and manage to keep up with their colleagues in L1-taught programs, who mainly have to focus on subject content. Learning in a second language slows down teaching and learning of subject content, as the CLIL subject teacher is curbed by the limited second language proficiency of the students (Oattes, Oostdam, De Graaff, & Wilschut, 2018). Does the assumed language barrier evoke a tailored pedagogical content knowledge (PCK) response from the CLIL teacher? This question adds to the empirical knowledge of how subject teachers deploy PCK to teach history in both mainstream and CLIL classes.

1.1. Pedagogical content knowledge

Teachers apply pedagogical content knowledge to select, present, connect and explain subject content suited to the specific characteristics and abilities of students (i.e., age, prior knowledge, and attainment level). Therefore, knowing when and how to apply pedagogical content knowledge is an important teachers' competence (Tuithof, 2017).

The conscious use of PCK is even more important in a CLIL context, where teaching content in an additional language creates new challenges. Bilingual education students in grade 7 have a limited vocabulary which makes it more difficult for the teacher to verbally explain historical concepts. Perhaps there is a need for a tailor-made CLIL PCK with 'tools' like extra illustrations, gestures, facial expressions, translanguaging, sounds, accentuation, charts, artefacts etcetera to bridge the language gap.

The integrated PCK concept was originally introduced by Shulman (1987). His initial idea was that teacher knowledge consisted of subject matter content knowledge, pedagogical content knowledge, and curricular knowledge. He defined PCK as: 'the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, represented, and adapted to the diverse interests and abilities of learners, and presented for instruction' (p. 8). This implies that the teacher has the essential skills to use appropriate representations and instructional strategies, and knowledge of students' understanding including language skills. PCK developed further, especially in science education, when three new theoretical elements were added by Magnusson, Krajcik, and Borko (1999): the teacher's knowledge of assessment, of the curriculum, and of teaching orientation. When we move from the general theory of PCK to the specific history PCK, the innovative study of Monte-Sano and Budano (2013) is essential. They identified four relevant PCK components: representing history, transforming history, attending to students' ideas about history, and framing history.

In this study into the impact of the language of instruction on history teachers' application of PCK, our approach is closest to Monte-Sano's and Budano's PCK component *framing history,* which is defined by how history teachers select and arrange topics into a coherent story that conveys cause-effect relations between and among events as well as the historical significance of events and people. In so doing they conceptualise and frame the history curriculum to illustrate significance, connections and interrelationships.

The current digital revolution has created new possibilities for teaching and learning, but education still depends on language, the most important tool at the disposal of the teacher. The realisation that content teaching is closely linked to language teaching has had consequences for PCK. Subject teachers have to be aware, careful and selective with the words they use, and adjust their PCK to the students' average language proficiency.

1.2. Bilingual education in the Netherlands

In the last 25 years BE in the Netherlands has become popular, as currently approximately twenty per cent of junior secondary schools decided to add an additional, parallel bilingual education department. Students in the junior schools (grades 7, 8 and 9) can choose to enroll in the mainstream or the bilingual education stream. The main goal of the BE-stream is for students to learn meaningful and content-rich English as a foreign language and to develop a sense of world citizenship (Nuffic, 2019). The latter is not part of this study.

At least fifty per cent of the Dutch BE curriculum in grades 7, 8 and 9 is offered in English, thus involving approximately five subjects and subject teachers plus the English teacher in CLIL. History is usually part of the BE curriculum.[2] Being a CLIL subject teacher implies teaching subject knowledge in a second language and simultaneously actively support the understanding, development and use of students' L2 proficiency.

CLIL pedagogy in the Netherlands is based on Westhoff's (2004) model of second language acquisition (SLA), which proposes five SLA-elements to be integrated in content teaching through a

[1] BE and CLIL are often used interchangeably. In this study, BE is referred to an organisational principle of a school, while CLIL refers to the pedagogical teaching principle to bring subject content and English together in CLIL classroom teaching (Mearns & De Graaff, 2018; Morton & Llinares, 2017).

[2] For more specifics on BE in the Netherlands see De Graaff & Van Wilgenburg, 2015.

second language. CLIL pedagogy contains students' abundant exposure to spoken and written language, e.g. when a new topic is introduced. The teacher will organize learning activities on meaning focused processing to facilitate students understanding the new subject content. Additionally, form focused processing may be facilitated where subject content and second language vocabulary and grammar are explicitly linked. In order to learn and master the newly learned content and language, students are stimulated to produce communicative output, by speaking and or writing (group) assignments. Moreover, the use of communication strategies is stimulated, to support students compensating gaps in their proficiency of the second language.

Both subject teachers and teachers of English are key players in deploying CLIL (Dale, Oostdam & Verschoor, 2018). However, different studies show that CLIL subject teachers prefer to focus on subject teaching and that they are, generally speaking, reluctant to engage in L2 teaching activities (Dale, Oostdam, & Verspoor, 2018; Koopman, Skeet, & de Graaff, 2014; Lorenzo, 2007); Oattes, Oostdam, De Graaff & Wilschut, 2018b).

The Dutch BE context offers a unique opportunity to compare and reveal differences in the PCK application of history teachers when teaching in Dutch and English because in the Netherlands parallel history lessons are taught in mainstream (MS) and BE stream in the same school, usually by the same history teacher, using the same curriculum, often using the same (literally translated) text and exercise books, during the same period of the year.

1.3. History and language

The history curriculum in junior secondary education focuses on developing a context-providing chronological knowledge framework to support skills that enhance historical consciousness. Students are taught a framework consisting of Ten Eras, each with four or five characteristic features, for example: *Era of Monks and Knights (500–1000)* with characteristic features like *'feudalism', 'manorial system and serfdom'*, and *'spread of Christianity in Europe'*. Understanding and applying organising ideas, or second order concepts, like causation, change, evidence, and significance are concepts of a different nature which requires students to use analytical and evaluative skills that will help them develop historical consciousness.

History entails the use of an extensive range of historical concepts, and therefore teaching concepts is essential to history lessons (Wilschut, Van Straaten, & Van Riessen, 2013). For many students with a limited language proficiency the academic school and subject-specific language used in history textbooks, assignments and teacher explanations is problematic. Students are known to struggle, even in their mother tongue, with tasks that require the use of general academic, subject-specific words, and formal language (Van Boxtel & Grever, 2011; Van Drie, Braaksma, & Van Boxtel, 2015; Wilschut, 2015). Explaining historical concepts in English to Dutch students with a limited English vocabulary thus becomes an extra challenge (Llinares & Morton, 2010; Lorenzo, 2013; Schleppegrell & De Oliveira, 2006). Dutch history teachers in BE grade 7, the first grade of junior secondary education, confirm this observation (Oattes, Oostdam, De Graaff, & Wilschut, 2018). Much teaching time in grade 7 has to be spent on additional explaining of information, translation, and contextualisation of 'difficult' texts (Oattes, Oostdam, De Graaff, & Wilschut, 2018). For example: explaining *serfdom* requires the expansion of the student's vocabulary and conceptual knowledge, as serfdom would be a new concept for them in Dutch as well. The history teacher uses appropriate subject-specific language (e.g. lord, peasant, serf, domain, manor, corvée), and related general academic language (e.g. dependency, obligation, obedience, protection, hierarchy, freedom), which may also contain unknown concepts and words.

1.4. The present study

To date studies are mostly focusing on science oriented PCK, or on CLIL (cf. Nikula, Dalton-Puffer, & Llinares, 2013). This small scale, explorative study focuses on the possible educational consequences on the application of Dutch teachers' PCK in a second language. Will the students' limited English proficiency force teachers to adjust their PCK? The teachers' language use is important, and therefore we recommend a specific linguistic study into PCK as a topic for future research. We expect to contribute to the knowledge bases of both classroom CLIL subject teaching and the analysis of applied PCK in two languages. We therefore adopted Shulman's original concept of PCK and Monte-Sano and Budano's (2013) specific PCK component *framing history,* by comparing history teachers' classroom instruction of subject content in Dutch and English.

Our research question is: *To what extent is the teacher's application of pedagogical content knowledge affected when history lessons in Dutch junior secondary bilingual education are taught in English?*

2. Method

2.1. Sample

The sample consisted of twelve digitally recorded sets of paired lessons (with a range of 50–80 min), each set consisting of one Dutch and one English spoken lesson about the same historical topic taught by the same history teacher to both a MS class and a BE class of the same secondary school. The focus of this study was on the teachers' instructional part of the lessons, because we wanted to establish if teachers' PCK-approach and explanation of historical concepts was affected by the language of instruction they used. The lessons were taught at seven different schools across the country. Four paired lessons were recorded in grade 7, and eight paired lessons in grade 9 of the pre-university education stream, involving approximately 200 junior students from 7th grade and 400 9th graders, between the ages of 12–15. The BE students are known to be extra motivated, high capacity learners (Mearns, de Graaff & Doyle, 2017) who can keep up, or even outperform their mainstream counterparts, certainly in grade 9 (Oattes, Fukkink, Oostdam, De Graaff, & Wilschut, 2020; Verspoor, De Bot, & Xu, 2015).

The seven teachers involved in this study (Teachers A- G) were regular history teachers whose native language was Dutch. Three teachers had a bachelor's degree and four had a master's degree in history education. They were educated and certified in using English as the language of instruction and communication in the CLIL classroom. All teachers had a minimally required Cambridge Advanced English (CAE) certificate, and four of them also had the Cambridge Proficiency English (CPE) certificate or similar qualification. They also all passed CLIL-training courses, either through in-school training or through external institutes. Their proficiency in CLIL and English is kept up to date by means of yearly refresher courses that are usually supervised by the English departments.[3]

[3] In an earlier study (Oattes, Oostdam, De Graaff, Fukkink, & Wilschut, 2018) a sample of 8 Dutch CLIL teachers were assessed on their spoken production and spoken interaction by English language experts using the In-Service Certificate in English Language Teaching (ICELT) marking scheme for Spoken English (CELA, 2015). The overall scores (min-max: 1–4 (*SD*) were 3.49 (*0.56*) for Spoken production and 3.31 (*0.73*) for Spoken interaction.

The history teachers, two females and five males, had between 8 and 29 years ($M = 15.14$, $SD = 7.43$) of history teaching experience and between 6 and 10 years ($M = 8.00$, $SD = 1.41$) of CLIL teaching experience. One of the teachers (A) taught in grades 7 and 9, the others taught either in grade 7 or grade 9 (see Table 3). Six teachers used the same, literally translated, history text and exercise book in both streams. One teacher used an English-language history textbook specially developed for Dutch bilingual education in grade 7, and an American history textbook focusing on world history with many of the 20th century topics that are mentioned in the Dutch curriculum in grade 9. The participating CLIL-teachers used the history textbook as the foundation of their curriculum-based history lessons. In grades 7, 8, and 9 the Ten Eras and their specific characteristic features form a chronologic narrative and are mainly taught through textbook and teacher driven instruction. In grades 10, 11, and 12 the Ten Eras are revisited, in the mother tongue, often with a more diachronic approach and a focus on developing students' historical knowledge, and their historical thinking and reasoning through challenging assignments, that may include evaluating and judging, taking a well-founded stand etcetera.

All observed lessons were presented by teachers using digital visual support, usually containing lesson goals, chronological information, multimodal visualizations of historical events, key concepts and assignments (homework). The recorded lessons were part of the curriculum and were selected by the history teacher, which establishes the ecological validity of the sample. Further, the lessons were authentic, typical, and most teachers used whole class instruction, often a combination of teacher's lecturing and teacher-student dialogic discussion, to introduce, position, and explain new historical concepts.

2.2. Measurements

We employed three instruments to collect data: a quantitative PCK scoring list to register the frequency and quality with which certain PCK elements were employed. To complement the quantitative findings, we used qualitative data software to analyse fourteen key elements from twelve paired lessons to distinguish differences and similarities in the language of instruction. Furthermore, we used a close reading technique (Fang, 2016; Greenham, 2019; Snow & O'Connor, 2016) to analyse and evaluate the content at word level of two key concepts as they were taught by the history teacher in both languages. Together with a limited linguistic analysis of the teachers' explanations of the concepts these instruments provided both a broad and a detailed perspective on the history teachers' PCK application in a CLIL context.

2.3. The quantitative PCK-observation instrument

For a quantitative analysis of the history teachers' application of PCK, the existing observation models of teaching behavior are geared towards mainstream education and appeared to be too generally pedagogic-didactic (Van de Grift, 2007), too language-oriented (De Graaff, Koopman, Anikina & Westhof, 2007), or too intellectually demanding for the involved junior students. PCK elements like Bloom's (1956) higher order thinking skills (analysing, evaluating, creating; see Krahtwohl, 2002), historical perspective taking (Huijgen, 2018), using current events or expressing personal opinions (King, Newman & Carmichael, 2009; Saye et al. (2018) could not be part of the observation model as they were hardly used.

We noticed that the majority of the observed teachers were too occupied with introducing a basic (grade 7) or a more detailed (grade 9) chronological narrative with the related facts and basic key concepts. Several observed history teachers had difficulties reaching a sufficient level of teaching (a minimum score of 15 out of the maximum of 30 points), because of a lack of clarity in the organisation, structure and/or execution of the lesson.

We therefore decided to use an inductive approach and developed and implemented our own instrument with a focus on a limited number of PCK elements that were actually displayed by the history teachers and underpinned by key elements of Shulman's (1987) PCK description. We concretized them for analysing history teachers' classroom instruction using Monte-Sano and Budano's (2013) PCK-category *framing history*. Next, we used the Protocol for Assessing the Teaching of History (PATH) designed by van Hover, Hicks, and Cotton (2012), an observation scheme including six categories with the goal to improve instruction. Elements of three categories, Lesson Components, Narratives, and Interpretations, fit in with our ideas on teachers' application of PCK. Also Gautschi (2015) who subdivided 'good history teaching' in three categories: lesson content, lesson structure and lesson impact, using 15 indicators. We slightly adapted three indicators to fit in our observation scheme, e.g. 'addressing changes over time and the connection between developments', 'factual accuracy, multiple perspectives, controversy', and 'awareness of historical witnesses and changes over time'. The observation instrument eventually consisted of four analytic PCK categories and a holistic category that were labelled: *Core of the lesson, Concepts, Judgement & Perspective, Chronology & Context,* and *Holistic Assessment.*

Core of the lesson focuses on the history teacher's capability to use details (poignant and illustrative examples) to support and strengthen the core of the lesson without these details becoming dominant or weakening the core of the lesson. The 'big picture' or the main theme of the lesson remains the central focal point (Gautschi, 2015; Hattie, 2012; van Hover et al., 2012). *Concepts* is aimed at the history teacher's command of the subject knowledge, the ability to concretise, provide current and appropriate examples, and give alternative descriptions (Gautschi, 2015; van Hover et al., 2012).

The next two categories include a more demanding in-depth approach of teaching and understanding of history concepts. *Judgement & Perspective* emphasises the exemplary role of the history teacher, who should make a clear distinction between an opinion and a judgment. The teacher does not give or invite unfounded opinion(s), but stimulates students to judge, take a position, and clarify one or several perspectives (Gautschi, 2015; Stradling, 2010; van Hover et al., 2012) on an evaluative question based on historical arguments. The teacher emphasises the role of coincidence in historical development (van Straaten, Claassen, Groot, Raven & Wilschut, 2015). *Chronology & Context* refers to the history teacher's capability to apply chronology and different concepts of time (era and dates), and places events or peoples' actions in the correct historical context without getting distracted (Huijgen, 2018; van Hover et al., 2012). *Holistic Assessment* is a complementing category to appraise the teacher's overall performance more spontaneously and impressionistically, as proposed by Sadler (2009), to be a counter weight to breaking down the observation in analytic categories and using strictly defined pre-set assessment criteria (see Appendix).

Experts used the quantitative PCK observation instrument that included four rating rubrics (see Table 1 for an example) integrated in a seven-point Likert scale (min-max: 0–6), that has four benchmarks: (a) insufficiently recognisable (0 = insufficient, 1: just short of sufficient); (b) partly recognisable (2 = just sufficient, 3: sufficient); (c] recognisable and comprehensible (4 = more than

Table 1
Description of a rating rubric of one analytic category of the PCK observation instrument.

Category 2: Concepts	
Score = 0 'insufficient'	The teacher uses no/insufficient/inaccurate description of key concept(s).
Score = 2 'just sufficient'	The teacher provides almost complete and accurate content or definition of key concept(s) without using examples.
Score = 4 'more than sufficient'	The teacher provides accurate content (definition) of key concept(s) using appropriate examples. Or: teacher disproves misconception(s).
Score = 6 'very good'	The teacher provides multiple alternative descriptions of key concept(s) and uses functional examples to explain, clarify.

Note. If score 2 does not do justice to the PCK demonstrated by the teacher, but score 4 is not (fully) achieved, then score 3 offers a middle course. (see Appendix for the complete instrument).

Table 2
Correlations of PCK categories.

PCK variable	1	2	3	4	5
1. Core of the lesson	–				
2. Concepts	.709**	–			
3. Judgement & Perspective	.464*	.516*	–		
4. Chronology & Context	.224	.109	.340	–	
5. Holistic Assessment	.766**	.731**	.651**	.393	–

Note. **. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed.

sufficient, 5: good); (d) comprehensible and providing insight (6 = very good). The interval sum score of the five individual scores (min-max: 0–30) made up the Lesson Score (see Table 3).

The observation instrument provided sufficient distinctiveness when rating the teachers' application of different PCK-categories (see Table 3, Lesson Score), and the five distinct categories reached sufficient interrater agreement: *Core of the lesson* .89; *Concepts* .89; *Judgement & Perspective* .88; *Chronology & Context* 0.72, and *Holistic Assessment* 0.86 (intra class correlation coefficient (ICC), two-way random, consistency).

We calculated correlations to examine the interrelationship between the five categories in the instrument (See Table 2). This analysis of bivariate correlations showed statistically significant correlations, varying from a low .109 to a high 0.766 Three of the four analytic categories (1, 2 and 3) were significantly related; only category 4, *Chronology & Context,* showed no significant relationship with the other three scales or with the Holistic Assessment category. The other three analytical PCK categories also proved to be significant predictors of the *Holistic Assessment.*

2.4. The qualitative PCK analysis instruments

We have added complementary qualitative analyses of the 'concepts' category to bring the teacher's voice back to life and to gain a deeper understanding of the teacher's crucial oral explanations as regards the quantitative analysis.

Firstly, the lesson transcripts of instructional teacher talk were used for open and axial coding, to label and select fourteen key elements of the paired lessons, and for further analysis using MaxQDA (Verbi Software, 2018). We supposed that a pattern could become visible as regards the comparison between the teaching language and the number of words used in the explanation (Blumenstock, 2008). We compared the teachers' oral explanations in both languages of fourteen diverse and representative key elements (e.g. *concepts, processes, learning activities)* that reflected the core of the history lessons (see Table 4). We only focused on the teachers' historical explanations and left out all other utterances, i.e. regulative remarks, unrelated teacher-student interaction. Using word processing software, we could determine: a) the total number of words used by the teacher in Dutch and English to explain the key element, b) the number of words in corresponding episodes, where the explanation in English and Dutch focused on similar examples and/or content details, and c) the number of words in unique episodes, where Dutch and English spoken explanations differed as a result of the teachers' choice to add (or leave out) a sub-topic or an extra content detail for clarification of the key element.

Secondly, we selected two transcribed teacher explanations of historical key concepts based on comparable content and size to zoom in on both explanations using a close content reading strategy. Originally rooted in the literary tradition, we used close reading aimed at enhancing literacy in readers through a focus on reading comprehension of complex content texts (Fang, 2016; Snow & O'Connor, 2016). By reading and rereading the selected fragments (see Examples 1 and 2) we determined the exact meaning of the explanations. This analysis was followed by a semantic and syntactic comparison and analysis (Greenham, 2019) to demonstrate similarities and differences in history teachers' linguistic presentation of the selected historical concepts in both languages and in both grades.

2.5. Procedure

All involved CLIL teachers volunteered and consented to participate in this study on the basis of anonymity after an appeal in a questionnaire on bilingual education, designed by the first author for a previous study (Oattes, Oostdam, De Graaff, & Wilschut, 2018). In 2015–2016 the first author made all video and sound recordings on location. Paired lessons taught by the same history teacher were usually recorded within the same week.

The 24 selected history lessons were analysed by seven experts (three history teacher educators and four experienced history teachers). Each lesson was analysed by three experts: the first author, as expert who analysed all the lessons, plus a random selection of two assessors from a group of six experts. To limit the burden of time investment, each expert assessed eight lessons, equally divided and randomly chosen from mainstream and bilingual grades 7 and 9. In order to prevent contamination, the experts were allowed to assess only one of the paired lessons.

The experts received a training and practiced using the observation and rating instrument in two formal sessions. Firstly, each expert individually watched one Dutch spoken and one English spoken recorded lesson without further instruction to get an impression of the material they were going to assess. We then organized a formal joint session to introduce and practice with the observation and assessment instrument. In the subsequent discussion it became apparent that there were differences in the assessment of the observed teachers. After a debate on key elements of a history lesson, we limited the number of categories as certain PCK elements, such as historical perspective taking and linking past events with current affairs, were observed only once.

For the qualitative analyses we selected concept explanations from teachers with contrasting PCK performances: Teacher B (grade 7) had an above-average PCK lesson score, and Teacher G (grade 9) who had a below-average PCK lesson score. They represent the diversity within the sample of history teachers. Two researchers independently checked and double-checked the close content reading analysis of the teachers' concept explanations and reached full agreement on the results of the content analysis.

2.6. Data analysis

For the quantitative part, assessment scores of the 24 observed lessons were analysed using SPSS (IBM, version 25) to calculate descriptive statistics. In addition, we determined effect-sizes to quantify differences between the application of the selected PCK categories in Dutch or English spoken lessons.

We checked the total number of words used in the teachers' explanations of a key element in both languages. Unique episodes were marked in both the Dutch and English lesson and the words used were counted. The total number of words used in unique episodes was subtracted from the word total to determine the number of words in corresponding episodes.

The close content reading analysis of the two historical concepts were carried out in the original languages, but for the reader's convenience the recorded explanations in Dutch were translated into English (see Examples 1 and 2). We added some linguistic characteristics of both concept explanations to establish similarities in total number of words, sentences, and words per sentence. We also compared Type-Token Ratio (TTR), or the vocabulary variation that is found in a spoken or written text by dividing the number of unique words by the total number of words. Finally, we looked at the lexical density (LD) of the explanations, that displays the linguistic complexity of a text based on functional and content words.

3. Results

3.1. Quantitative results

Table 3 shows the scores of the individual teachers (A-G) and lessons (1–12), and the mean scores of the PCK categories of both grades and both streams. Overall, there was little variation between the history teachers' PCK application in mainstream (Dutch) and (English) CLIL lessons. Individual mainstream and CLIL results between paired lessons in both grades never differed more than 2.5 points (min-max = 0–30), while the mean lesson scores were nearly identical These average scores (approximately 15 out of 30) suggest that the application of PCK in both languages and both grades can be improved. Overall, the magnitude of the differences in our study are small-to-medium for the four PCK categories and the *Holistic Assessment* for lessons from both grades 7 and grades 9 (see the overview presented in Table 3).

3.1.1. PCK in grades 7

The ratings indicate that 7th grade history teachers are focused on teaching historical knowledge through teaching historical concepts. Teachers likely assume that the 7th graders are not yet ready to be taught a more in-depth approach like historical thinking and reasoning through teaching *Judgement & Perspective* and

Table 3
PCK Category assessment scores. Mean (standard deviation) and effect size (d) of history teachers' application of pedagogical content knowledge. within grades 7 and 9 (d_{BE-MS}).

	Teacher	Core of the lesson		Concepts		Judgement & Perspective		Chronology & Context		Holistic Assessment		Lesson Score	
		MS	BE	MS	BE	MS	BE	MS	BE	MS	BE	MS	BE
Grade 7													
Lesson 1	A	4.50	5.00	3.50	5.00	3.00	2.50	2.50	2.00	4.00	4.00	17.5	18.5
Lesson 2	B	4.00	5.00	4.00	4.00	2.00	2.50	1.50	2.00	3.50	4.00	15.0	17.5
Lesson 3	B	5.00	4.50	4.50	3.50	2.00	2.00	1.00	1.50	3.00	3.00	15.5	14.5
Lesson 4	C	4.00	4.00	3.00	3.00	1.00	1.00	3.00	2.00	3.00	2.00	14.0	12.0
Mean		**4.38**	**4.63**	**3.75**	**3.88**	**2.00**	**2.00**	**2.00**	**1.88**	**3.38**	**3.25**	**15.5**	**15.6**
(SD)		*(0.48)*	*(0.48)*	*(0.65)*	*(0.85)*	*(0.82)*	*(0.71)*	*(0.91)*	*(0.25)*	*(0.48)*	*(0.96)*	*(1.47)*	*(2.95)*
d_{BE-MS}		0.52		0.17		0		−0.18		−0.17		0.04	
Grade 9													
Lesson 5	D	3.50	1.50	3.00	3.50	1.50	2.50	2.00	2.50	3.33	3.00	13.3	13.0
Lesson 6	D	3.00	2.00	3.67	3.00	2.00	2.00	1.50	2.00	2.00	2.50	12.2	11.5
Lesson 7	A	5.50	5.00	4.00	4.00	4.00	4.00	3.00	3.00	4.00	4.33	20.5	20.3
Lesson 8	E	4.50	4.00	3.00	4.50	3.00	3.00	2.50	2.50	4.00	4.00	17.0	16.0
Lesson 9	E	5.00	4.50	4.00	3.50	2.00	2.50	4.00	4.00	4.50	4.00	19.5	18.5
Lesson 10	F	4.00	3.00	3.00	3.00	2.00	2.50	3.00	2.50	2.50	3.00	14.5	14.0
Lesson 11	G	2.00	1.50	0.33	1.00	1.00	1.50	1.00	2.50	2.00	1.33	6.3	7.8
Lesson 12	G	2.00	3.00	2.00	1.50	2.00	2.00	2.50	2.50	1.50	2.00	10.0	11.0
Mean		**3.69**	**3.06**	**2.88**	**3.00**	**2.19**	**2.50**	**2.44**	**2.69**	**2.98**	**3.02**	**14.2**	**14.1**
(SD)		*(1.31)*	*(1.35)*	*(1.22)*	*(1.20)*	*(0.92)*	*(0.76)*	*(0.94)*	*(0.59)*	*(1.12)*	*(1.06)*	*(4.79)*	*(4.13)*
d_{BE-MS}		−0.47		0.10		0.37		0.32		0.04		0.02	

Note. MS = mainstream, BE = bilingual education stream, d_{BE-MS} = Cohen's *d (difference between BE-score compared to MS score)*, Mean score PCK categories: min-max: 0–6 for analytic and holistic grading, Lesson Score: min-max = 0–30.

Chronology & Context. In grade 7, both BE and MS, the history teachers scored highest on *Core of the lesson,* while *Concepts* scored second best. In comparison *Judgement & Perspective* and *Chronology & Context* received substantially lower scores. The *Holistic Assessment* of the MS and BE lessons was almost identical. Teachers were focusing on *Core of the lesson* more strongly in BE than in MS (medium-sized $d = 0.52$). Teachers' use of *Concepts, Judgement & Perspective,* and *Chronology & Context* received similar scores in both streams.

3.1.2. PCK in grades 9

The ratings indicate that history teachers in grade 9 apply a broader form of PCK, where all four analytic categories are part of the history lesson. Teaching concepts in grade 9 seems to be more strongly linked to historical thinking and reasoning. In both grade 9 streams the history teachers scored highest on *Core of the lesson* closely followed by *Concepts*. However, compared to grade 7 the scores have decreased. *Judgement & Perspectives* and *Chronology & Context* received lower scores, but compared to grade 7 both category scores increased. The *Holistic Assessment* of the teacher performance in both grade 9 streams was almost identical. A closer look at the results show that teachers scored substantially less on *Core of the lesson* in BE than in MS (medium-sized $d = -0.47$). The comparison *Chronology & Context* in BE and MS showed an opposite effect, as the BE teachers scored slightly higher on this category (small-sized $d = 0.32$). Comparing the teachers' scores on *Judgement & Perspective* also scored slightly higher in BE than MS (small-sized $d = 0.37$). Scores for *Concepts* and *Holistic Assessment* in BE and MS were similar.

While the mean lesson scores in grades 7 and in grades 9 showed a strong similarity, the overall individual lesson score (min-max: 0–30) showed substantial differences. In grade 7 the highest individual overall lesson score in MS was 17.5 and the lowest was 14.0. In BE the highest score was 18.5 and the lowest was 12.0. In grade 9 the individual scores were 20.5–6.3 in MS and 20.3–7.8 in BE. Considering that the maximum score was 30, these results suggest that, even when we exclude the outlier, most teachers have room for improving their PCK in the analysed categories in both languages.

3.2. Qualitative results

We first performed a word count of the 14 selected key elements of the twelve paired lessons. The word total in Table 4 showed that teachers in grade 7 used more words in English (±20%) than in Dutch for the explanations, while in grade 9 it was the other way round. On average the differences in words used in corresponding episodes were limited: ± 6% more Dutch words in grade 7, and ± 5% more English words in grade 9. This limited difference indicates that the teacher used very similar explanations in both languages, for instance by using the same, but translated, explanation. Explanations of key elements in grade 9 differed between 20% and 50% in the number of words in corresponding episodes. This indicated more diversity in the teachers' choice of words used to explain the key elements in the different languages.

To establish if there was a substantive difference between the teacher's explanations in both languages, we compared two explanations of two key concepts using close content reading to compare the historical content. We analysed two (transcribed) teachers' oral explanations in Dutch and English: *medieval domain* in grade 7 (see Example 1), and *guerrilla-warfare* in grade 9 (see Example 2).

Table 4
Overview of teachers' words spent on key elements (Word Total), and percentages of words used in Corresponding Episodes of teachers'. explanations in MS and BE in grades 7 and 9.

Key elements of the lesson		Teacher	Word Total		Used words in Episode	
			MS	BE	MS	BE
Grade 7						
1	Historical thinking game: Odd one out	A	353	383	97.7%	88.5%
2	Christianisation: Nature gods	B	452	448	62.8%	78.1%
3a	Medieval domains: Third estate	B	257	308	75.1%	39.3%
3b	Religion: Origins of Islam	B	814	784	80.1%	84.7%
4	Middle-Ages: Writing assignment	C	395	559	79.5%	71.4%
Mean			**454**	**577**	**79.0%**	**72.4%**
Grade 9						
5a	Nazi-Germany: Antisemitism	A	362	255	65.5%	69.8%
5b	Nazi-Germany: Holocaust	A	280	381	96.8%	65.6%
6	Interbellum: Hitler's rise to power	D	835	402	37.0%	85.3%
7	World War 1: Versailles Treaty, 1919	D	267	386	83.5%	53.4%
8	World War 1: Submarine warfare	E	587	889	100%	52.6%
9	World War 1: Black Hand	E	556	214	52.2%	100%
10	Interbellum: Hitler and Sudetenland	F	445	186	63.8%	100%
11	Vietnam War: Lyrics assignment	G	257	296	93.4%	85.1%
12	Vietnam War: Chronological overview	G	204	235	42.2%	65.1%
Mean			**421**	**360**	**70.5%**	**75.2%**

Example 1.
MS and BE grade 7: the *medieval domain* according to Teacher B (lesson 3).

Box 1a
Excerpt from lesson 3; Literal translation of the Dutch version.

> This domain is how the people in the Middle Ages lived. So, there were no big cities. There were mainly small villages. Small villages, like this one, and they nearly all had the same structure. You can also see my drawing here on the black board. You can see the manor, or sometimes the lord's castle. That is where the lord of the domain lived and he was also a knight. He actually protected the village and it was sometimes built on a hill.
>
> You should have also drawn it yourself on your blueprint. Next to it is a church. Christianity was important in the Middle Ages. Christian faith spreads, and nowadays you can see in villages that the church is the oldest building and also often located in the Centre of the village. Sometimes this was the only stone building, thereby showing us how important faith really was. Around it are small farms, as you can see on this drawing, where the ordinary farmers lived and worked the land. Sometimes near open water, sometimes near a river. Often there were woods close by where the pigs could range freely, and where wood could be gathered for the fireplace. And of course, there was farming land. The farmers partly had to work the lord's farmland, and partly their own land. That was what you had to draw, because we will continue with this subject in this paragraph.

Box 1b
Excerpt from lesson 3; the English version.

Last time we had this drawing of a domain, you might remember. So on the domain, you had to draw certain places, certain things. You had to for example draw the house of a lord, which is called a manor or even a castle sometimes. You also had to draw a church. If you'll visit some villages in the Netherlands, and also in other countries, you will see that many villages have a church as a centre building. So, Christianity was spread and then the church becomes an important building in the village. And sometimes it was even, sometimes together with the manor, it was the only building made of stone. Because you can see that the farms that surround the church are made of wood. You also made farmlands, and you made a river and a mill that the farmers could use, but they had to pay taxes for it to the lord, so it was actually owned by the lord. The lord of this domain was also a knight and this knight had to protect the farmers. So, you could find out that these farmers lived close to his castle or manor because he could protect them. Because there were times that there were wars or even Vikings who invaded the coast and they destroyed whole villages. And if you don't live close to someone who can protect you, it can cost you your life.

The Dutch and English explanations of Example 1 consisted of an equal number of words, but in Dutch more sentences were used: 19 to 13. English sentences contained more words: 18.38 to 12.58. In Dutch the TTR (51.48%–46.41%) and the LD (39.33%–35.98%) were higher.

Both in Dutch and in English the description of a medieval domain was the core of Teacher B's explanation, in this case connected to an earlier drawing assignment. The words used by Teacher B were quite similar as both explanations included *manor, castle, church, Christianity, village centre, stone building, protection, taxes, farmland, knight.* However, on closer inspection some differences were apparent.

The English explanation contained three elements that were not mentioned in the Dutch version: farms were made of wood, the presence of a mill and using it would cost you, and the protection offered, for instance, against Viking raids. The Dutch explanation contained six elements that were not mentioned in the English version: there were only a few big cities, the castle was often positioned on a hill, the church nowadays is often the oldest building of a village, only important buildings were made of stone, description of the village's location, the distinction between land of the lord and land of the farmers, and the farmers' obligation to work the lord's land.

Teacher B kept both explanations on the same track and included many similar words. Therefore, both explanations resembled each other and both were useful, but they were not identical. The Dutch explanation came out on top because it contained more relevant historical details that elucidated the overall picture of the key concept *medieval domain* best.

Example 2.

MS and BE grade 9: *guerrilla-warfare in Vietnam* according to Teacher G (lesson 12).

Box 2a
Excerpt from lesson 12; Literal translation of the Dutch version.

Who doesn't know what a guerrilla war is? Okay. I always call a guerrilla war a sort of war of resistance, with people from the resistance fighting the war. It is not two armies against each other, no, it's a sort of secret war that launches attacks on army troops, that's possible. Or you go and steal their food supplies, or you blow up an army base, or a pub full of American soldiers. Then we call it a guerrilla war. So, many of the terrorists -the word terrorist is quite difficult to position when you think about it- are involved in guerrilla wars. They feel like soldiers, but are not wearing green outfits and then start using their Kalashnikovs to fight other people. No, it happens more like under the surface. Why is the word terrorist emotionally charged? Everyone sees himself as a freedom fighter, the IS warriors also believe that, but we are the victims, so they are the terrorists. But in fact, they are fighting a guerrilla war, certainly here in the West. So, then the guerrilla war starts in South Vietnam, where there are people opposing their government. So, the North-Vietnamese find support in South Vietnam for their war of resistance.

Box 2b
Excerpt from lesson 12; the English version.

Who doesn't know what a guerrilla war is? All right, I wrote it down somewhere in here … A guerrilla war is when you have not official soldiers in green suits but more like the resistance in the Netherlands in World War 1, 2 sorry, were fighting a guerrilla war. It's a war under the radar. The way I see it you're not official warfare but more 'hit and run' kind of actions. So, in the Vietnam situation it was North Vietnamese warriors infiltrating in South Vietnam and not everyone in South Vietnam was agreeing with their governments –a lot of corruption going on- so resistance fighters in South-Vietnam started a guerrilla war against their government, supported by the North Vietnamese. So, this was the war that America was getting into, a guerrilla war. The Vietnamese know Vietnam, they know the ins and outs of Vietnam. This is a problem. The Americans are new, it's not their country, it's not their climate. It's a bit like the German army fighting the Soviet Union in the Second World War when they were not really ready to fight in winter. Here we have got a similar problem and that makes it scary. The Americans feel like in a way they're fighting ghosts.

The Dutch and English explanations of Example 2 were almost equal in size: 209 to 217, but in Dutch a few more sentences were used: 14 to 12. English sentences contained more words: 14.93 to 18.08. In Dutch the TTR was higher (55.34%–49.53%), while the LD (39.33%–35.98%) was almost equal (47.37%–48.85%).

Guerrilla-warfare was at the core of these explanations, and resistance was a key word used in both, but the context and examples differed quite a lot between the versions. An attempt was made to define the concept, but Teacher G did not have a clear-cut definition at hand *(… I wrote it down somewhere in here …)* and this led to two different explanations.

The English explanation mentions *'not official soldiers in green suits'* and later on *'not official warfare'*, and *'under the radar'*. The Dutch version states: *'no two armies against each other. No, it was more of a war of resistance', 'a secret war"*. Without further elaboration, Teacher G used the terms *'under the radar'* and *'hit-and-run kind of actions'*, which raises the question whether students understood the meaning and whether these terms clarified or confused the explanation. In the Dutch version, on the other hand, Teacher G gave four clear examples of guerrilla-warfare.

Only in the Dutch version did Teacher G try to connect Vietnamese guerrilla-warfare with current IS terrorism. In the English version the teacher elaborated on Vietnam as a difficult battleground for the Americans and made a comparison with German troops in the Russian winters in World War 2. Without any further elaboration the explanation finishes with *'the Americans feel like in a way they are fighting ghosts'*. This remark was not connected to any part of the given explanation. Teacher G had difficulty defining and keeping close to the original concept in both languages, and wandered off the track on several occasions. Both explanations were incomplete and seemed improvised, but eventually the Dutch explanation of the concept *guerrilla-warfare* contained more relevant historical details than the English version.

4. Discussion and conclusion

The research question driving this study was to what extent the application of history teachers' PCK is affected when history is taught in a second language (English) in Dutch junior secondary bilingual education schools.

Summarizing, the main conclusion is that there was little difference between the teachers' application of PCK when history is taught in English or in Dutch both in grades 7 and in grades 9. Contrary to what we expected CLIL teachers did not apply a tailor-made PCK for their bilingual education students. We did find differences in the teachers' application of PCK between grades 7 and grades 9. Further, history teachers in grade 9 performed better in the PCK categories *Judgement & Perspective* and *Chronology & Context* than 7th grade teachers. The quantitative results also showed that the mean PCK scores of the history teachers were average (approximately 15 points out of a possible 30), neither high nor low, again in both grades.

Complementary qualitative results confirmed that, overall, CLIL teachers in grade 7 spent the majority of their words on the explanation of key concepts, while in grade 9 this was substantially less. The word-for-word content analysis of explanations of two historical key concepts in two languages showed teachers using slightly more relevant historical details in L1. A limited linguistic analysis established small differences between the teacher's use of L1 and L2.

4.1. Difference in the application of PCK in grades 7 and 9

Teaching history lessons in English in grade 7 is no simple matter for CLIL teachers, as their students' English proficiency is limited and they can only use a modest, but developing, vocabulary (Coonan, 2007; Moate, 2011; Oattes, Oostdam, De Graaff, Fukkink, & Wilschut, 2018; Papaja, 2013). This language issue seemed to lead CLIL history teachers to focus strongly on the PCK categories *Core of the lesson* and *Concepts*. We expected, but did not observe, history teachers to use an adjusted PCK to compensate for the students' limited English proficiency. However, the teachers' PCK focus in Dutch was on the same two categories, even though there was not a language issue that slowed down the teaching process. One could expect that history teachers in Dutch in grade 7 would use the advantage of the mother tongue proficiency to step up and introduce new subject content and/or challenging assignments. But we determined that in both streams, the PCK categories *Judgement & Perspective,* and *Chronology & Context* were hardly involved, hence their lower scores. In terms of *Holistic Assessment,* no significant difference was observed between teachers' application of PCK and the language used.

In grade 9 BE teaching circumstances were different as 9th graders have developed a (much) better English proficiency and possess more comprehensive historical knowledge and skills which they built up in grades 7, 8, and 9. According to CLIL teachers this makes teaching in English more attractive (Coonan, 2007; Moate, 2011; Oattes, Oostdam, De Graaff, & Wilschut, 2018; Papaja, 2013). In 9th grade BE, interaction has become increasingly natural and more prominent (from the CLIL perspective of stimulating students' language output), enabling history teachers' PCK focus to gradually shift from a teacher-driven to a more teacher-student-driven teaching classroom (Nikula et al., 2013). We observed a shift towards a more advanced PCK focus in both streams, with less emphasis on *Core of the lesson* and *Concepts,* and higher scores for the more in-depth history categories *Judgement & Perspective* and *Chronology & Context*. In terms of *Holistic Assessment*, we found that the history teachers in BE and MS grade 9 were rated equally, suggesting that their use of PCK was of similar quality regardless of the language used.

Overall, we conclude that the English teaching language used in grades 7 and 9 does not significantly influence the teachers' application of PCK. The English and Dutch spoken history lessons were subjected to the same teaching approach. Presumably, teachers preferred to keep both streams and grades on the same track, at the same pace and using the same PCK for organisational motifs (Oattes, Oostdam, De Graaff, & Wilschut, 2018) thereby ignoring the distinction between the two language groups.

4.2. Average PCK performance

The mean Lesson Score of the five PCK categories in grade 7 (min. 0-max. 30) was very similar in both streams, while the language proficiency in Dutch and English was not. This language advantage of the MS was not reflected in the results. No individual teacher differed more than 2.5 points between their PCK score in MS and BE, which suggested that teaching in English or Dutch hardly affected their application of PCK. The mean lesson score of the five PCK categories in grade 9 (min. 0-max. 30) did not improve compared to grade 7 and was nearly identical in both streams. It is striking to see that individual teachers differed no more than 1.5 points between their MS and BE lesson scores, which suggests that also in grade 9 teaching in English or Dutch led to similar teacher application of PCK.

We conclude that there is a strong similarity in the teachers' application of PCK regardless of the teaching language that is used. Based on this research it seems that teachers design one lesson plan, which is both practical and reassuring for them in terms of teaching similar historical content, and then teach it in both English and Dutch. Also noted is that the mean lesson scores were average with approximately 15 out of the max. 30 points. This suggests that there is room for history teachers to improve their knowledge and skills as regards the analysed PCK categories, irrespective of the language (L1 or L2) in which they teach.

4.3. Choice of words

The quantitative outcome of the applied PCK as mentioned above was complemented with two qualitative evaluations of the teachers' language use in explaining fourteen key elements in Dutch and English. Results from the word count showed that in grades 7 four out of five explanations of historical key elements shared approximately 85% of the content, thus indicating that the content of the paired lessons was quite similar. The word count qualitative evaluation in grade 9 showed that MS and BE explanations of nine historical key elements of the lessons, in 7 cases, shared between 50 and 80% of the content, indicating that the explanation of content of the paired lessons was less similar than in grade 7.

These outcomes were not unexpected in view of the fact that BE students' limited foreign language proficiency in grade 7 enhances passive language use (listening and reading) and curbs teacher-student interaction, thereby leaving most of the communication up to the teacher. 9th Graders possess a higher proficiency in English which allows them to participate more actively in the CLIL lessons and thus influence the content of the lesson, for instance with spontaneous questions or unscripted discussions.

Generally speaking the Dutch and English explanations of the concepts in both grades did not differ that much linguistically. It was obvious though, that compared to grade 7, the explanations in grade 9 had a higher vocabulary variation (TTR) and a higher linguistic complexity (LD) in both languages. When we closely examined the explanations of a historical key concept in both grades and in both languages, we found that the explanations in grades 7 and 9 were not perfect nor identical. The English versions contained some historical details that were lacking in the Dutch version, and vice versa. Ultimately, both Dutch versions contained more relevant historical detail than the English versions. These two examples suggest that the history teacher was slightly more proficient using the mother tongue. Other concept explanations and/or other teachers may lead to different results.

4.4. Limitations and suggestions for future research

To appraise the results of this explorative, small scale study into this new domain we need to be aware of a number of limitations that may have affected its outcome. Nevertheless, we want to share our findings as a stimulus for further research into the role of the subject teachers' PCK application in CLIL.

First of all, we realise that the low sample size prohibited the use of inferential statistics which means that the described results are suggestive.

Secondly, we focused our analyses only on the instructional part of the lesson, in particular on teachers' explanations of historical concepts. Other elements of the lesson, for instance the teachers' interaction in the CLIL classroom, were not analysed and could lead to additional insight (cf. Dalton-Puffer, 2007; Nikula et al., 2013).

Thirdly, we have looked into a small number of PCK categories with a new observation instrument that can be retested and adjusted. In hindsight it may have been better to split up the two categories *Judgement & Perspective* and *Chronology & Context* in four separate categories. *Chronology* was an integral part of all the observed lessons, while *Context* was clearly not. This influenced the correlation with the other analytic categories and the holistic assessment. Future research may want to enlarge the sample, extend the observation phases and add observation categories, to get a more complete picture of how a CLIL subject teacher manages his CLIL lessons.

Fourthly, we used qualitative analyses of teachers' explanations concerning historical content in both Dutch and English as an illustration to the abstract word count analysis. The analysis of two key concepts also included a lexical exploration. An extended comparative syntactic analysis could also reveal lexical diversity in the teachers' language use and quality of the taught subject content knowledge in the different streams. (cf. Dallinger, Jonkman, Hollm, & Fiege, 2016; Gablasova, 2010).

Finally, in this study most CLIL history teachers used a textbook and exercise book literally translated from Dutch, which likely encouraged them to develop identical teaching materials. u Using a different history textbook and exercise book for the CLIL lessons could lead to more PCK-diversity and more custom-made teaching (materials). Or for that matter, it would be interesting to find out if history teachers who are also English native speakers apply PCK in the same manner as the non-native history teachers.

4.5. Final remarks

There is little urgency for BE teachers to change their uniform teaching approach as BE students are known to be extra motivated, high capacity learners (Mearns, de Graaff & Doyle, 2017). But BE students, especially those in grade 7, lag behind their mainstream counterparts in historical knowledge due to their L2 deficiency (Oattes et al., 2020; Verspoor et al., 2015). This study fits in with earlier research into the role of the CLIL subject teacher (cf. Cammarata & Tedick, 2012; Coonan, 2007; Moate, 2011; Oattes, Oostdam, De Graaff, & Wilschut, 2018; Papaja, 2013) that suggested that keeping both language streams on the same track and at the same pace, using the same PCK approach, is convenient for CLIL teachers for organisational reasons. But is this 'one size fits all' the most suitable approach for students in the BE stream? History teachers easily flip the language switch when they enter CLIL classrooms, but they do not seem to consider to develop and teach a tailor made PCK approach best suited to the needs of the BE students.

Appendix

PCK Analytic and Holistic Observation, Analysis, and Rating Instrument.

Rubrics to score four main analytic categories and a holistic category of pedagogical content knowledge (PCK) as demonstrated by the history teacher in class.

Category 1: Core of the lesson	
Goal: teacher holds on to the core (theme, subject, big picture - whether or not captured in the goals of the lesson) of the lesson and uses details (poignant and illustrative examples) that support and enhance the core of the lesson.	
Example 1	*Core: role and importance of the clergy in the Middle Ages.* *Details: daily schedule of monastic life, distinction between higher and lower clergy.*
Example 2	*Core: course of the Second World War 1939–1945.* *Details: zoom in on start, turning points, end of the war, using concrete cases.*

(*continued*)

Category 1: Core of the lesson	
Score = 0 'insufficient'	Teacher confuses core of lesson and details; does not distinguish between core and details; uses incorrect details; teacher strays (is distracted) from core, resulting in core of lesson not becoming clear/apparent.
Score = 2 'just sufficient'	Teacher makes core of lesson recognisable for students, but uses few or no details.
Score = 4 'more than sufficient'	Teacher makes core of lesson comprehensible and supports this with (multiple) details. The 'what?' questions are discussed.
Score = 6 'very good'	Teacher makes core of lesson comprehensible and provides insight, using a rich variety of powerful details. The 'why?' questions are discussed.
Scores 1, 3 and 5	If score 2 does not do justice to the PCK demonstrated by the teacher, but score 4 is not (fully) achieved, then score 3 offers a solution.
Category 2: Concepts	
Goal: The teacher uses the key concept (a key concept is the basis of a subject) in a skilful manner (in command of the knowledge, able to concretise, provides suitable and topical example(s)) and is able to provide alternative descriptions.	
Example	*Key concept: autarky = economic self-sufficiency.* *This key concept has different appearances in different eras and/or different locations.*
Score = 0 'insufficient'	The teacher uses no/insufficient/inaccurate description of key concept(s).
Score = 2 'just sufficient'	Teacher provides almost complete and accurate content or definition of key concept(s) without using examples.
Score = 4 'more than sufficient'	Teacher provides accurate content (definition) of key concept(s) using appropriate examples. Or: teacher disproves misconception(s).
Score = 6 'very good'	Teacher provides multiple alternative descriptions of key concept(s) and uses functional examples to explain, clarify.
Scores 1, 3 and 5	If score 2 does not do justice to the PCK demonstrated by the teacher, but score 4 is not (fully) achieved, then score 3 offers a solution.
Category 3: Judgement & Perspective	
Goal: the teacher gives an opinion, a judgment, takes a position, chooses one or more perspectives on an evaluative question/statement and substantiates his choice with historical arguments. He assigns a role to coincidence in history (bonus).	
Opinion = (here) subjective point of view, insufficiently supported by substantive arguments.	
Judgment = balanced, well-founded view	
Perspective = looking and judging from a particular context	
Coincidence = unpredictable, unintentional, undirected	
Example	Could the Second World War have been prevented? Yes, if it had not been for Hitler, none of it would have happened (opinion).
Example	You can answer that with either yes or no. Yes, if you look at the following factors … …/No, if you look at the following factors … … (judgment).
Example	If you look at it from a German point of view, then … … From an English point of view, however, it could have … …, or: If you look at it from the perspective of the economic situation in the 1930s, you can better understand why people … … (perspective).
Example	These events occurred (un)intentionally, but it could also have gone/ended differently (not common in history methods, therefore only included in score 5 and score 6).
Score = 0 'insufficient'	Teacher judges in a presentist manner/imposes a one-sided judgment on students/gives an opinion without historical arguments.
Score = 2 'just sufficient'	Teacher follows the perspective of the method, does not take personal position/perspective; history presented as a self-evident (series of) development(s).
Score = 4 'more than sufficient'	The teacher presents a perspective that is recognisable for the students and provides historical arguments/judgments based on historical grounds.
Score = 6 'very good'	The teacher makes different, correctly argued judgments/points of view/perspectives recognisable for the students/attention for the role of coincidence in history (historical discussion, clear examples of multi-perspectivity).
Scores 1, 3 and 5	If score 2 does not do justice to the PCK demonstrated by the teacher, but score 4 is not (fully) achieved, then score 3 offers a solution.
Category 4: Chronology & Context	
Goal: the teacher places the subject in the different concepts of time and in the appropriate context of that time, so that students get a better picture and gain deeper understanding of this other (less familiar) time.	
Chronology = Working with historical time (years, eras, periods).	
Context = the main historical characteristics of a particular period.	
Example chronology	*Around 1500 Western-European explorers left in search of an alternative sea route to Asia. This is the era of Explorers and Reformers, between 1500 and 1600. It concludes the Middle Ages and marks the start of modern history.*
Example context	*Around 1500 Western-Europe is in a period of transition, during which medieval society slowly changed as a result of the voyages of discovery, changes in religion and in urban societies, among other things.*
Score = 0 'insufficient'	The teacher has not placed the subject, or placed it incorrectly in time (period/dates) or links an incorrect context to that time.
Score = 2 'just sufficient'	The teacher uses simple dating of subject (period/dates) without placing it in a context/without outlining a portrait of the era.
Score = 4 'more than sufficient'	The teacher links the subject to extensive dating (years, eras, periods) and outlines a suitable context for that time.
Score = 6 'very good'	The teacher discusses the subject in combination with the complexity of the time (f.e. the synchronicity of the asynchronous) and/or encourages the students to use chronological reasoning, in which the correct context of the time is designated.
Scores 1, 3 and 5	If score 2 does not do justice to the PCK demonstrated by the teacher, but score 4 is not (fully) achieved, then score 3 offers a solution.
Category 5: Holistic Assessment	
Goal: rating of the history teacher's overall use of PCK. This category is not composed of the sum of the four analytical PCK category scale scores, but of the overall PCK performance as observed by the assessor.	
Scores = 0 - 6	from insufficient (0, 1) to (just, more than) sufficient (2, 3, 4) to (very) good (5, 6).

References

Bloom, B. (1956). *Taxonomy of educational Objectives. Handbook I: The Cognitive domain*. New York: David McKay Co Inc.

Blumenstock, J. (2008). Size matters: Word count as a measure of quality on Wikipedia. April 21-25 April. In *[paper presentation]: WWW '08: Proceedings of the 17th international conference on world Wide Web*. https://doi.org/10.1145/1367497.1367673. Beijing, China.

Cammarata, L., & Tedick, D. (2012). Balancing content and language in instruction: The experience of immersion teachers. *The Modern Language Journal, 96*(2), 251–269.

Cenoz, J. (2015). Content-based instruction and content and Language integrated learning: The same or different? *Language Culture and Curriculum, 28*(1), 8–24. https://doi.org/10.1080/07908318.1000922

Coonan, C. (2007). Insider Views of the CLIL class through teacher self observation-introspection. *International Journal of Bilingual Education and Bilingualism, 10*(5), 625–646.

Coyle, D., Hood, P., & Marsh, D. (2010). *Content and language integrated learning*. Cambridge, United Kingdom: Cambridge University Press.

Dale, L., Oostdam, R. J., & Verspoor, M. (2018). Towards a professional development tool or teachers of English in bilingual streams: The dynamics of beliefs and practices. *International Journal of Bilingual Education and Bilingualism*. https://

doi.org/10.1080/13670050.2018.1556244
Dallinger, S., Jonkman, K., Hollm, J., & Fiege, C. (2016). The effect of content and language integrated learning on students' English and history competences. Killing two birds with one stone? *Learning and Instruction, 41*, 23–31.
Dalton-Puffer, C. (2007). *Discourse in Content and Language Integrated learning (CLIL) classrooms*. Amsterdam, The Netherlands: John Benjamins Publishing. https://doi.org/10.1075/lllt.20
Dalton-Puffer, C., & Nikula. (2014). Content and language integrated learning. *Language Learning Journal, 42*(2), 117–122.
Dalton-Puffer, C., & Smit, U. (2013). Content-and-Language Integrated learning: A research agenda. *Language Teaching, 46*(4), 545–559.
De Graaff, R., Koopman, G., Anikina, Y., & Westhoff, G. (2007). An observation tool for effective L2 pedagogy in Content and Language Integrated Learning (CLIL). *International Journal of Bilingual Education and Bilingualism, 10*(5), 603–624. https://doi.org/10.2167/BEB462.0
De Graaff, R., & Van Wilgenburg, O. (2015). The Netherlands: Quality control as a driving force in bilingual education. In P. Mehisto, & F. Genesee (Eds.), *Building bilingual education systems: Forces, mechanisms and counterweights* (pp. 167–179). Cambridge, United Kingdom: Cambridge University Press.
European Commission. (2009). EU language policy. Retrieved from http://ec.europa.eu/education/languages/eu-language-policy/index_en.htm.
European Commission/EACEA/Eurydice. (2017). *Key Data on teaching languages at school in Europe – 2017 edition*. Luxembourg: Publishing office of the European Union. Eurydice Report.
Fang, Z. (2016). Teaching close reading with complex texts across content areas. *Research in the Teaching of English, 51*(1), 106–116.
Gablasova, D. (2014). Issues in the assessment of bilingual educated students: Expressing subject knowledge trough L1 and L2. *Language Learning Journal, 4*(2), 151–164.
Gautschi, P. (2015). Good history teaching in Switzerland : Principles, findings, suggestions. In A. Chapman, & A. Wilschut (Eds.), *Joined up history: New directions in history education research* (pp. 137–160). Charlotte, NC: Information Age Publishing.
Greenham, D. (2019). *Close reading. The basics*. London, UK/New York, NY: Routledge (Taylor & Francis Group.
Hattie, J. (2012). *Visual learning for teachers: Maximizing impact on learning*. Milton, United Kingdom: Routledge.
van Hover, S., Hicks, D., & Cotton, S. (2012). Can you make 'Historiography sound more friendly?": Towards the construction of a reliable and validated history teaching observation instrument. *The History Teacher, 45*(4), 603–611.
Huijgen, T. (2018). *Balancing between the present and the past. Promoting students' ability to perform contextualization (doctoral dissertation)*. Groningen, The Netherlands: University of Groningen. https://doi.org/10.1080/09571736.2014.889974
Jaffee, A. (2016). Social studies pedagogy for latino/a newcomer youth: Toward a theory of culturally and linguistically relevant citizenship education. *Theory & Research in Social Education, 44*(2), 147–183.
King, M., Newmann, F., & Carmichael, D. (2009). Authentic intellectual work: Common standards for teaching social studies. *Social Education, 73*(1), 43–49.
Koopman, G-J, Skeet, J, & de Graaff, R (2014). Exploring content teachers' knowledge of language pedagogy: a report on a small-scale research project in a Dutch CLIL context. *The Language Learning Journal, 42*(2), 123–136. https://doi.org/10.1080/09571736.2014.889974
Krahtwohl, D. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice, 41*(4), 212–218.
Lin, A. M. Y. (2016). *Language across the curriculum and CLIL in English-as-an-additional-language contexts: Theory and practice*. Dordrecht: Springer.
Llinares, A., & Morton, T. (2010). Historical explanations as situated practice in content and language integrated learning. *Classroom Discourse, 1*(1), 46–65.
Lorenzo, F. (2007). An analytical framework of language integration in L2 content based courses: The European dimension. *Language and Education, 21*, 503–516. https://doi.org/10.2167/le708.0
Lorenzo, F. (2013). Genre-based curricula: Multilingual academic literacy in content and Language integrated learning. *International Journal of Bilingual Education and Bilingualism, 16*(3), 375–388.
Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome, & N. Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 95–132). Dordrecht, the Netherlands: Kluwer.
Mearns, T., & De Graaff, R. (2018). Bilingual education and CLIL in The Netherlands: The paradigm and the pedagogy. *Dutch Journal of Applied Linguistics, 7*(2), 122–128. https://doi.org/10.1075/dujal.00002.int
Moate, J. (2011). The impact of foreign language mediated teaching on teachers' sense of professional integrity in the CLIL classroom. *European Journal Of Teacher Education, 34*(3), 333–346.
Monte-Sano, C., & Budano, C. (2013). Developing and enacting pedagogical content knowledge for teaching history: An exploration of two novice teachers' growth over three years. *The Journal of the Learning Sciences, 22*(2), 171–211.
Morton, T., & Llinares, A. (2017). Content and language integrated learning (CLIL): Type of programme or pedagogical model? In A. Llinares, & T. Morton (Eds.), *Applied linguistics perspectives on CLIL* (pp. 1–16). Amsterdam, the Netherlands: John Benjamins Publishing Company.
Nikula, T., Dalton-Puffer, C., & Llinares, A. (2013). CLIL classroom discourse: Research from Europe. *Journal of Immersion and Content-Based Language Education, 1*(1), 70–100. https://doi.org/10.1075/jicb.1.1.04nik
Nuffic. (2019). *Kwaliteitsstandaard tweetalig onderwijs 2.0* [Standard for bilingual education 2.0] Retrieved from https://www.nuffic.nl/sites/default/files/2020-8/kwaliteitsstandaard-tweetalig-onderwijs2-0.pdf.
Oattes, H., Fukkink, R., Oostdam, R., De Graaff, R., & Wilschut, A. (2020). A showdown between bilingual and mainstream education: The impact of language of instruction on learning subject content knowledge. *International Journal of Bilingual Education and Bilingualism*. https://doi.org/10.1080/13670050.2020.1718592. Advance online publication 31 Jan 2020.
Oattes, H., Oostdam, R., De Graaff, R., Fukkink, R., & Wilschut, A. (2018b). CLIL in Dutch bilingual education: How Dutch history teachers focus on second language learning. *Dutch Journal of Applied Linguistics, 7*(2), 156–176, 10.1075/dujal.18003.oat.
Oattes, H., Oostdam, R., De Graaff, R., & Wilschut, A. (2018a). The challenge of balancing content and language: Perceptions of Dutch bilingual education history teachers. *Teaching and Teacher Education, 70*, 165–174, 10.1016./j.tate.2017.11.022.
Papaja, K. (2013). The role of a teacher in a CLIL classroom. *Glottodidactica, 40*, 147–154.
Pérez-Cañado, M. (2012). CLIL research in europe: Past, present, and future. *International Journal Of Bilingual Education And Bilingualism, 15*(3), 315–341.
Sadler, D. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. In G. Joughin (Ed.), *(2009). Assessment, learning and Judgement in higher education* (pp. 45–63). Dordrecht, the Netherlands: Springer Science. https://doi.org/10.1007/978-1-4020-8905-3_4.
Saye, J., Stoddard, J., Gerwin, D., Libresco, A., Maddox, E., & The Social Studies Inquiry Research Collaborative (SSIRC). (2018). Authentic pedagogy: Examining intellectual challenge in social studies classrooms. *Journal of Curriculum Studies, 50*(6), 865–884. https://doi.org/10.1080/00220272.2018.1473496
Schleppegrell, M., & De Oliveira, L. (2006). An integrated language and content approach for history teachers. *Journal of English for Academic Purposes, 5*, 254–268. https://doi.org/10.1016/j.jjeap.2006.08.003
Sclafani, C. (2017). Strategies for educators of bilingual students: A critical review of literature. *International Journal of Education and Literacy Studies, 5*(2), 1–9. https://doi.org/10.7575/aiac.ijels.v.5n.2p.1
Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–22. https://doi.org/10.17763/haer.57.1.j463w79r56455411
Snow, C., & O'Connor, C. (2016). Close reading and far-reaching classroom discussion: Fostering a vital connection. *Journal of Education, 196*(1), 1–8.
Stradling, R. (2010). *Multiperspectivity in history teaching: A guide for teachers*. Strasbourg, France: Council of Europe, Publishing division, Directorate of communication and research.
Tucker, G. (2001). *A global perspective on bilingualism and bilingual education*. Georgetown University Round Table on Languages and Linguistics. Retrieved from: https://moam.info/a-global-perspective-on-bilingualism-and-bilingual-education-pdf-_599ad6da1723dd09401acd74.html.
Tuithof, H. (2017). *The characteristiscs of Dutch experiences history teachers' PCK in the context of a curriculum innovation (doctoral dissertation)*. Utrecht, The Netherlands: Utrecht University.
Van Boxtel, C., & Grever, M. (2011). Between disenchantment and high expectations. History education in The Netherlands, 1968-2008. In E. Erdmann, & W. Hasberg (Eds.), *Facing, mapping, bridging diversity: Foundation of a European discourse on history education* (Vol. 2, pp. 83–116). Germany: Wochenschau Verlag. Schwalbach/Ts.
Van Drie, J., Braaksma, M., & Van Boxtel, C. (2015). Writing in History: Effects of writing instruction on historical reasoning and text quality. *Journal of Writing Research, 7*(1), 123–156.
Van de Grift, W. (2007). Quality of teaching in four European countries: A review of the literature and an application of an assessment instrument. *Educational Research, 49*(2), 127–152. https://doi.org/10.1080/00131880701369651
Verbi Software. (2019). *MAXQDA 2020 [computer software]*. Berlin, Germany: VERBI Software (Available from: maxqda.com).
Verspoor, M., De Bot, K., & Xu, X. (2015). The effects of English bilingual education in The Netherlands. *Journal of Immersion and Content Based Language Education, 3*(1), 4–27.
Westhoff, G. (2004). The art of playing a pinball machine. Characteristics of effective SLA- tasks. *Babylonia, 12*, 58–63.
Wilschut, A. (2015). Testing frame of reference knowledge in national examinations. Report on an Experiment in The Netherlands. In A. Chapman, & A. Wilschut (Eds.), *Joined up history: New directions in history education research* (pp. 85–114). Charlotte, NC: Information Age Publishing.
Wilschut, A., Van Straaten, T., & Van Riessen. (2013). *Geschiedenisdidactiek: Handboek voor de vakdocent* (2nd ed.). Bussum, The Netherlands: Coutinho.