Data Article

Multicultural classroom discourse dataset on teachers' and students' dialogic empathy

Check for updates

Chrysi Rapanta [a,*], Cláudia Gonçalves [a], João Rui Pereira [a], Dilar Cascalheira [a], Beatriz Gil [a], Rita Morais [a], Anna Čermáková [b], Julia Peck [b], Benjamin Brummernhenrich [c], Regina Jucks [c], Mercè Garcia-Milà [d], Andrea Miralda-Banda [d], José Luna [d], Maria Vrikki [e], Maria Evagorou [e], Fabrizio Macagno [a]

[a] Faculty of Social Sciences and Humanities, Universidade NOVA de Lisboa, Avenida de Berna 26, 1069-061, Lisbon, Portugal
[b] Faculty of Education, University of Cambridge, UK
[c] Institute of Psychology for Education, University of Münster, Germany
[d] Cognition, Development and Educational Psychology Department, University of Barcelona, Spain
[e] Department of Education, University of Nicosia, Cyprus

ARTICLE INFO

ABSTRACT

This dataset belongs to a large-scale educational intervention project introducing cultural literacy learning in schools. The dataset includes transcribed dialogic interactions from 111 lessons conducted in classrooms of five countries: England, Portugal, Germany, Spain, and Cyprus. The lessons were part of the same Cultural Literacy Learning Programme made available to teachers from three age groups (pre-primary, primary, secondary) to implement with their ordinary classes during ten sessions. The data are from the third and eighth sessions and followed the same structure, content, and objectives. As the main goal of the project to which the dataset belongs was on cultural literacy learning, understood as dialogic dispositions and values, the data were coded according to a Dialogic Empathy coding scheme presented in detail in Macagno et al. (2020). This rich multicultural, and multilinguistic, dataset is offered for further analysis by dif-

* Corresponding author.
  E-mail address: crapanta@fcsh.unl.pt (C. Rapanta).
  Social media: (C. Rapanta)

ferent types of researchers, such as linguists interested in intercultural pragmatics or educational psychologists interested in cross-sectional studies of dialogue and reasoning skills.

Specifications Table

Subject	Social Sciences
Specific subject area	Classroom discourse
Type of data	Dataset
How data were acquired	Through audio-recording of classroom interactions using at least two professional microphones and two professional recorders per class, as well as non-participant classroom observation and note-taking techniques.
Data format	Cleaned and labelled raw data accompanied by a separate annotation.
Parameters for data collection	Data were collected after an open call for voluntary participation for schools/teachers was launched by each project partner. All participant teachers followed professional development aiming at the implementation of the same lesson plans, forming part of a pedagogical programme designed by the research team in collaboration with a piloting teachers' team.
Description of data collection	The classroom sessions were recorded by research team members with audio recorders. In some cases, video recording was used as well to facilitate transcription, especially when it comes to speakers' recognition issues in large classes. The transcriptions were made by the researchers, who assisted the classes taking notes, and the translations by the researchers or externally contracted translators.
Data source location	Institution: University of Cambridge City/Town/Region: several Country: England Institution: Universidade NOVA de Lisboa City/Town/Region: several Country: Portugal Institution: University of Münster City/Town/Region: several Country: Germany Institution: University of Barcelona City/Town/Region: Catalunya Country: Spain Institution: University of Nicosia City/Town/Region: several (only in the Greek side of the island) Country: Cyprus
Data accessibility	With the article

Value of the Data

- The dataset is representative of how students from 5 to 15 years old and from different nationalities interact with each other negotiating their different cultural identities [2,3]. Culture here is not perceived exclusively as belonging to a national/ethnical/linguistic group but as expressing diverse viewpoints and trying to include the other in one's own discourse.
- The data come from classrooms in five different countries with students working on the same lesson plans, previously constructed as part of a multi-party European project. The variable of 'same lesson plan' gives a certain comparability to the data, as it allows researchers to look at how students with different cultural backgrounds (ethnical, linguistic, social, age-related, etc.) interact with the teacher and with each other on similar topics.
- The data are of high value also for researchers interested in dialogic teaching [4] as the same teacher/classroom is observed at two moments, once in the beginning (third session of the

Cultural Literacy Learning Programme) and once towards the end (eighth session of the Cultural Literacy Learning Programme).

- The data can form the object of further analyses within an intercultural pragmatics [5] perspective as well as from a cross-sectional analysis point of view (i.e. comparing between different age groups).

1. Data Description

This paper contains data from a multilingual corpus of a wider Horizon 2020 project called DIALLS (Dialogue and Argumentation for cultural Literacy Learning in Schools; for more information see www.dialls2020.eu). It consists of a dataset of transcripts of classroom interactions of students from ages 5 to 15 years old. The topic of the project and the classroom interactions is cultural literacy through dialogue and argumentation among school children. The project includes a Cultural Literacy Learning Programme (the CLLP) to teach children dialogue and argumentation skills whereby children can learn to engage critically and inclusively with each other's ideas about a series of themes around social responsibility, living together and 'being European' in the 21st century. For a description of the CLLP design and contents see: https://dialls2020.eu/wp-content/uploads/2019/06/D3.1-final.pdf.

The dataset includes transcripts of classroom interactions collected between September 2019 and March 2020 from 71 classrooms across Europe (see Table 1). The same lesson structures were implemented in every classroom, according to the age group, namely pre-primary, primary, and secondary school students. Each project partner selected the schools and classes to be recorded, transcribed, and translated. The classroom sessions were recorded by research team members in each project country with professional audio recorders and microphones (in some cases video recordings were also available to facilitate the audio record transcription). The transcriptions were produced in Excel (.xlsx) files following a pre-established structure and were made by the researchers who had also assisted the sessions in a non-participant observation mode. For the transcriptions, a simplified adaptation of Jefferson's [7] conventions was used (see Table 2). The translations were made by the researchers or externally contracted translators. Students and teachers' names were anonymized and identified as Student 1 (S1), Student 2 (S2), … or Teacher (T). Each partner was responsible for anonymizing their own data following General Data Protection Regulation procedures.

The learning programme used wordless short films and wordless picturebooks to stimulate classroom discussions and was designed for three age groups: age group A (pre-primary), age group B (primary) and age group C (secondary). The same programme was followed by the dif-

Table 1
Participant schools/classrooms demographic data.

	UK	Portugal	Germany	Spain	Cyprus
No of classrooms					
Pre-primary	6	5	1	5	10
Primary	5	2	5	4	6
Secondary	3	5	10	4	0
KL1 sessions					
Pre-primary	6	5	1	5	9
Primary	5	2	5	4	6
Secondary	3	4	10	2	0
KL2 sessions					
Pre-primary	6	4	1	4	4
Primary	5	2	1	4	0
Secondary	3	4	2	4	0
Total Sessions	28	21	20	23	19

Table 2
Simplified transcription conventions used in the dataset.

Symbol	Phenomenon
[…]	Long pauses (important also at the end of a turn if there is a gap before the next speaker speaks)
-	Abrupt cut
SPEech	Higher volume/emphasis of a (part of a) word or phrase.
[ ]	Overlapping speech (also colored in yellow in the dataset to further indicate this).
{}	Inaudible segment not transcribed or only partially transcribed or reconstructed.
(writes)	Description of a relevant nonverbal activity or relevant paralinguistic aspects that allow understanding of the activity (gasps, chatter sounds, etc.) or relevant indications that allow understanding of whom the speech is addressed to. *Example: Wait your turn please! (to S1) Go on (to S2).*

ferent participant classrooms, presented in Table 1. Two sessions implemented by a random selection of participating classes were recorded. The recorded classes are distinguished as follows:

Key-point Lesson 1 (KL1): Session number three of the Cultural Literacy Learning Programme. The session followed a dialogic lesson plan about the film 'Ant[1]' for age group A, 'Papa's boy[2]' for age group B, and the book 'Eccentric city[3]' for age group C.

Key-point Lesson 2 (KL2): Session number eight of the Cultural Literacy Learning Programme. This session was transversal to the three age groups, and it followed a dialogic lesson plan, adapted for each age, about the film 'Baboon on the Moon[4]'.

As it can be noted in Table 1, the numbers of KL1 and KL2 sessions do not correspond, meaning that not always it was possible for a classroom to implement both sessions. This was due to the Covid19 pandemic lockdown restrictions implemented already in March 2020, which was when the KL2 data were planned to be collected.

2. Design, Materials and Methods

2.1. Research goal

In this paper, we present data from a multi-country design-based research study leading to the development of an innovative, transversal and cross-curricular citizenship education programme. The study's main hypothesis is that students will improve their cultural literacy learning skills [6] through engaging with contents and practices that relate to cultural literacy values and dispositions (i.e., tolerance, empathy, inclusion, etc.). This article contains data from the following countries of the project consortium, namely UK, Portugal, Germany, Spain, and Cyprus.

2.2. Research design and methods

To identify whether students changed their cultural literacy learning skills during the implementation of the programme, we opted for collecting and transcribing classroom discourse data at two moments, namely session number three (identified in the dataset as KL1) and session number eight (identified as KL2). Cultural literacy was operationalized in terms of dialogic empathy, which was suggested by the researchers as an intermediary concept akin to grasp how

---

[1] Filmmaker: Julia Ocker (Germany, 2017).
[2] Filmmaker: Leevi Lemmetty (Finland/Ireland, 2010).
[3] Illustrator: Béatrice Coron (France, 2014).
[4] Filmmaker: Christopher Duriez (UK, 2002).

participants take or not the Other into consideration in their own discourse. Transcript data was then manually coded with the DIALLS coding scheme aiming at grasping students' and teachers' dialogic empathy, as explained in detail in Macagno et al. [1]. The inter-rater reliability among trained coders was calculated, reaching a satisfying result (Krippendorf's $a = .77$), and any discrepancies among the coders were resolved through discussions until final full agreement was reached. The DIALLS coding scheme provides ordinal categories (from 0 to 8) which indicate the degree of dialogicity in the classroom discourse ('0' stands for dialogically irrelevant moves). The moves 1-8 do not have the same dialogical level, being divided in low-dialogical and high-dialogical moves. Low dialogical moves are the following: Managerial (1), Stating (2), and Accepting/Discarding (3). High dialogical moves are the following: Expanding (4), Inviting (5), Meta-dialogical (6), Reasoning (7), and Meta-dialogical Reasoning (8). Each turn was coded using only one code.

The dataset is organised in an Excel file divided into several sheets, each sheet representing a country. Each country page contains randomly but organized by session (starting at KL1) the transcripts and translations of the implemented sessions (first the age group A, then B, then C). The first sheet in the dataset presents the index of all the sessions; the index has hyperlinks to the transcripts for easier navigation. Files are named following the structure: Country_Session number_Age group_Key-point Lesson_Classroom number (e.g., UK_1_A_KL1_C1 stands for a transcribed session curated by our UK partner of classroom No 1 of pre-primary school students implementing Key-point Lesson 1).

Ethics Statement

There were two ethics assessment levels for the project: 1) The local level where each country had to adhere to its own processes to acquire ethical clearance for the within-country data collection processes; and 2) a wider project level, secured by the lead partner (University of Cambridge), guaranteeing that opinions/approvals by ethics committees and/or competent authorities for the research with human subjects, and specifically with children, were obtained by all partners before the data collection. In particular, all data collection and treatment processes respected international ethical research standards related to informed consents granted by all participants (teachers and student caretakers) before the beginning of the project, guaranteeing their voluntary participation and personal data protection throughout the project. Individual students and groups of students were included in the audio and/or video recordings of the classes only when the previous consent by their parents/caretakers was obtained. Anonymization of both students and teachers was guaranteed since the moment of transcription, through the use of a password-protected master sheet containing the name of each participant and their identification code (e.g. S1, S2, etc.). Original audio and/or video files were saved in external discs, accessed by password, and safeguarded locally by each responsible institution.

CRediT Author Statement

Chrysi Rapanta: Conceptualization, Writing – original draft, Writing – review & editing, Data curation; **Cláudia Gonçalves:** Writing – original draft, Writing – review & editing; **João Rui Pereira:** Data curation, Data analysis, Writing – review & editing; **Dilar Cascalheira:** Data analysis; **Beatriz Gil:** Data analysis; **Rita Morais:** Data analysis; **Anna Čermáková:** Data analysis; **Julia Peck:** Data analysis; **Benjamin Brummernhenrich:** Data analysis; **Regina Jucks:** Data analysis; **Garcia-Milà:** Data analysis; **Andrea Miralda-Banda:** Data analysis; **JoséLuna:** Data analysis; **Maria Vrikki:** Data analysis; **Maria Evagorou:** Data analysis; **Fabrizio Macagno:** Data analysis, Supervision, Writing – review & editing.

Declaration of Competing Interest

No conflict of interest is declared by the authors.

Acknowledgments

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2021.107518.

References

[1] F. Macagno, C. Rapanta, E. Mayweg-Paus, M. Garcia-Mila, Coding empathy in dialogue, J. Pragmat. 2021. (submitted).
[2] S. Hall, P. Du Gay, Questions of cultural identity, Sage, London, 1996.
[3] C. Rapanta, M. Vrikki, M. Evagorou, Preparing culturally literate citizens through dialogue and argumentation: rethinking citizenship education, Curric. J. (2020), doi:10.1002/curj.95.
[4] N. Mercer, L. Dawes, J.K. Staarman, Dialogic teaching in the primary science classroom, Lang. Educ. 23 (2009) 353–369.
[5] I. Kecskes, Intercultural Pragmatics, Oxford University Press, Oxford, 2013.
[6] F. Maine, V. Cook, T. Lähdesmäki, Reconceptualizing cultural literacy as a dialogic practice, Lond. Rev. Educ. 17 (2019) 383–392.
[7] G. Jefferson, Glossary of transcript symbols with an introduction, in: G.H. Lerner (Ed.), Conversat. Anal. Stud. from First Gener., John Benjamins Publishing, Amsterdam, 2004, pp. 13–31.