

<p>Institution: University of Cambridge</p>
<p>Unit of Assessment: UoA28</p>
<p>Title of case study: The English Profile Programme and EF Research Unit</p>
<p>1. Summary of the impact (indicative maximum 100 words) The EPP Project identifies criterial features for second language acquisition. It has engaged stakeholders in the teaching and testing of language learners. This is facilitated by the EPP network and website. The project has enabled Cambridge Assessment to define the English language constructs underlying Cambridge examinations at different proficiency levels more explicitly. The work has improved the tests themselves, but also allowed Cambridge Assessment to better communicate the qualities of their tests for accreditation and recognition. Stakeholders are more actively engaged through provision of resources for teachers, testers, ministries of education etc., on the website, and in seminars. The project has led to further research with an international language school, which has led to teachers and parents of the school pupils being more aware of the needs for successful second language acquisition.</p>
<p>2. Underpinning research (indicative maximum 500 words) The English Profile Project (EPP) is a groundbreaking collaborative research program, registered with the Council of Europe, working to provide a detailed set of Reference Level Descriptors for English. These Reference Level Descriptors will provide concrete examples of the competencies laid out in the Common European Framework of Reference for Languages (CEFR), clearly describing what a learner of English can be expected to know at each level. The work is supported by data from the Cambridge Learner Corpus (CLC), consisting of exam scripts of students taking the Cambridge exams. The CLC is one of the largest learner corpora in the world, thereby bringing its own challenges in terms of possible use and analysis.</p> <p>The EPP project was initiated in 2007 in the University of Cambridge by Cambridge Assessment, the Computer Laboratory and the Research Centre for English and Applied Linguistics, now DTAL (Dept. of Theoretical and Applied Linguistics). The case study focusses on the contribution of members of DTAL. Central to the EPP in DTAL are: John Hawkins (Professor and Director of former Research Centre for English and Applied Linguistics, 2004-2013), Dora Alexopoulou (SRA, joined the department in 2008), Paula Buttery (Lecturer, joined 2006), Henriette Hendriks (Reader, joined 1998), and Teresa Parodi (Lecturer, joined 1997).</p> <p>The combination of staff gives substance to the theoretical dimension of the English Profile Project, especially in language acquisition theory and in the computational analysis of learner English at different stages: Hawkins is world-renowned in the area of linguistic complexity, research of importance for the project as part of teaching and assessing second language learning is related to understanding what linguistic phenomena are complex and therefore potentially difficult to acquire. Alexopoulou is a specialist in the syntax of language, and researches the second language data from that angle. Buttery is one of our computational linguists. Her interests are in the construction and evaluation of (psycho)-computational models of language acquisition, and in automated corpus analysis (grammatical and lexical acquisition from corpora). Hendriks has been working on first and second language acquisition research since the early 1990s and has coordinated one of the largest projects of its time on adult second language acquisition (the Structure of Learner Varieties, follow-up project from the ESF funded project Adult Immigrants' Second Language Acquisition (Klein and Perdue 1993)). Parodi has researched bilingual and second language acquisition since the early 1990s. Her specialty is in the UG approach to language acquisition.</p> <p>The EPP project led to further research co-operations, amongst others with an internationally based foreign language school, EF Education First, with whom Alexopoulou has been working to build an even larger learner corpus (40 million words and growing), the EFCambridge open language Database, that can be freely accessed for research purposes by all interested in second language acquisition. Whereas second language learner databases existed in the past (cf. the ESF corpus), they were always of a much smaller size, and mostly hand-coded. The big contribution Cambridge made in this project is to research automatic ways of up-to-date tagging and parsing</p>

Impact case study (REF3b)

the contents of the corpora, whereas previous corpora have often been searchable only at word-level. Other corpora of learner English do exist, but none is on the scale of the Cambridge Learner Corpus and the EFCamDAT corpus, and as far as we know, none has been tagged and parsed in the way we have done. Both corpora allow for the control of first language influences on second language learning, and the EFCamDAT corpus has the additional advantage that multiple learners with the same first language can be followed longitudinally across multiple years.

3. References to the research (indicative maximum of six references)

1. Alexopoulou, D. (2012). 'Automating Second Language Acquisition Research: Integrating Information Visualisation and Machine Learning'. In Proceedings of EACL, Joint Workshop of LINGVIS and UNCLH, Avignon, France, with H.Yannakoudakis (principal author) and T.Briscoe.
2. Alexopoulou T, H. Yannakoudakis and A. Salamoura, 2013 Classifying intermediate Learner English: a data driven approach to learner corpora, in S.Granger et al., *Twenty Years of Learner Corpus Research: Looking back, moving ahead*. Corpora and Language in Use, Proceedings I, Louvain-la-Neuve: Presse Universitaires de Louvain, pp.11-23.
3. Geertzen, J., Alexopoulou, T., Korhonen, A. (2013). Automatic Linguistics Annotation of Large Scale L2 Databases: The EF-Cambridge Open Language Database (EFCamDat). In: *Proceedings of the 31st SLRF*, Cascadilla Press.
4. Hawkins, J.A. & P. Buttery (2009). 'Using learner language from corpora to profile levels of proficiency: Insights from the English Profile Programme'. Proceedings of the 3rd ALTE Conference 2008, Cambridge University Press, Cambridge.
5. Hawkins, J.A. & L. Filipovic (2012) Criterial Features in the Learning of English: Specifying the Reference Levels of the Common European Framework. Cambridge University Press, Cambridge.

All outputs can be supplied by the University of Cambridge on request.

4. Details of the impact (indicative maximum 750 words)

Research output has led to improvements of assessment of language learning (through language test development and validation) and informing of teaching materials. Given that the project is set up in cooperation with Cambridge Assessment and Cambridge University Press, impact is worldwide, including students taking exams, teachers preparing the students for their exams, and more generally, teachers of English. Exam candidature has grown from 2 million in 2007 to over 4 million in 2013; something that Cambridge Assessment feel is partly attributable to the cooperation in the EPP project with researchers in DTAL [1]. Also, accreditation of the Cambridge suite of exams has been helped as the qualities of the tests can be better communicated to agencies such as OFQUAL, UKBS and equivalent overseas agencies such as DIAC in Australia and CIC in Canada based on the research by DTAL.

Impact of the project can also be measured through the participation in the EPP network events and the use of the EPP website. Both were set up to promote increased engagement of stakeholders (governments, teachers, language learners) with the research projects. The website informs of future and recent network events, has a link to the Cambridge Learner Corpus, but also provides useful resources for researchers, teachers, testers, ministries of education and other English Profile network partners. Visits on the website show active collaboration with practitioners. Resources include, amongst others, the *English Vocabulary Profile* and the *Guide to the CEFR for English Language Teachers* [2]. The former can be used by teachers, exam writers, materials developers and researchers to identify the words or phrases a learner can be expected to know at each level; to view words and phrases within a specific topic area; and to search for additional aspects of language, such as which uncountable nouns learners can be expected to know at A1, which verbs are frequently used in the passive at B2 and which words are used in which registers at different levels. The Vocabulary Profile is accompanied by webinars by the author (Annette Capel). The latter clarifies to teachers how the CEFR can be useful to them in terms of seeing what learners need to work on to attain a certain level; of creating their own assessment grids; and working out curriculum plans. The website has been up and running since approximately August 2008 and gets 1500-2000 visits per week. Since January 2010 there have been 138,928 unique visitors with 33,249 unique visitors since January 1 2013, of whom 70% are new and 30% are

returning visitors. Furthermore, currently 2496 people subscribed to the EPP newsletter, and 11,000 are registered users of English Vocabulary Profile.

An ever-growing number of government advisors and educationists make up the English Profile Network. Included in the non-academic collaborators are the Bosnia-Herzegovina Ministry of Education, Vietnamese National Institute for Educational Strategy, and the Bahrain's Bahraini Petroleum Company. Regular network meetings are held (twice per year) in which stakeholders are invited to be updated on the research, and interact with researchers on the application of results for their specific needs (for example, teaching plans, national curricula development, assessment and exam planning). Some of the stakeholders also participate as data collaborators to create a new database, the *Cambridge English Profile Corpus*. That is, schools (mostly secondary), language schools and universities provide data for the database. The schools are spread all over the world (Croatia, Argentina, Austria, the Netherlands, France, Russia, Vietnam), and have benefits when they participate. For example, they 1) gain on-line access to a subset of the English Profile Dataset (including the contributor's own data) in an accessible and easily searchable format. This can help the teachers understand their own students' needs better, and to develop teaching materials catering for their needs. For example, if they have been concentrating on the teaching of particular forms, they can measure how successful their teaching was (using the database), and if they find in the data that some forms are not actually acquired, they can adjust their teaching program accordingly. Presentations on how to use the database in the classroom for other purposes (show students examples of words with multiple meanings, collocations, etc.) are also available. 2) They receive free tickets to English Profile network workshops, which will further include training relevant to teachers, such as how to rate a student's work by CEFR level, and free access to the eBook versions of John Trim and Jan van Ek's Council of Europe volumes (the T-series: *Waystage*, *Threshold* and *Vantage*). 3) Schools finally also receive a 'certificate of participation', and listing of the school's name, with thanks, on the website's [corpus collection participants](#) page, thereby potentially improving / strengthening their profile.

Interest from language schools is also evident. EF Education First is one of the largest international education organisations in the world, with 400 offices and language schools, exchange programmes, and degree courses in over 50 countries world-wide. They teach students and professionals. EF Education First has funded a large research project in the Research Centre for English and Applied Linguistics, now DTAL (launch February 2010) and is now using 'big data' from a corpus jointly designed by EF Education First and the funded EF Research Unit at DTAL to further the understanding of criterial features related to CEFR language stages of fluency. This understanding of course also leads to improved learner directed teaching. EF Education First has furthermore aligned all their teaching with the CEFR, collaborating with framers of the EPP to do so. Recently (linked to their 20th anniversary teaching in China), it has also started to engage press, teachers and parents in issues regarding the best age to teach / learn a second language [3, 4, 5, 6], and the best ways to motivate learners. Work within the EF Research Unit feeds back directly into those issues, and the teaching material and assessment of language students' work [7].

5. Sources to corroborate the impact (indicative maximum of 10 references)

[1] Statement form person 2 (Director, Research and Validation; Cambridge English Language assessment)

[2] <http://www.englishprofile.org>

[3] <http://uk.prweb.com/releases/2013/9/prweb11085712.htm>

[4] <http://english.cri.cn/11354/2013/09/08/195s786621.htm>

[5] <http://www.best-news.us/news-5121497-EF-YOUTHS-English-Chinese-20th-anniversary-large-online-campaign-was-officially-launched.html>

[6] 13 EF-CN-EF in the News Document

[7] Person 1 (Vice-President for Academic Affairs; Education First) can be contacted for corroboration of this claim.