Presentation Abstracts & Presenters' Biographies

Contents

Not just different, but better

Nick Saville, Cambridge Assessment English

In 2020 ALTE celebrated 30 years as a multilingual association. The 7th International Conference planned for Madrid in April 2020 was intended as a celebration – looking back at the history, as well as looking forward to future prospects. However, the Covid-19 pandemic triggered a rapid rethink and change of plan that has led to new ways of working together. This year's digital symposium using the exVo platform is an example; it provides an opportunity for us to reflect on what has changed and is likely to be different from now on – as represented by the different sessions highlighted in the programme.

In education generally, emergency remote teaching was a necessary response to school closures, requiring pupils and their parents to engage in learning @home. This was enabled by digital tech – computers, tablets and broadband connection – and Zoom has changed the ways we interact in all walks of life. During the recovery from the initial Covid shock, we have started talking about the 'new normal', and online or hybrid events have replaced face-to-face and predominantly 'place-based' activities we were used to.

For high-stakes language assessment, @home testing with remote proctoring emerged as the only way to conduct assessments while traditional testing venues were closed. This proved to be successful as an emergency response, but also highlighted many legitimate concerns, including the ways in which AI is used to enable these new systems.

While there have been many positive elements in this rapid shift in social practice, existing challenges to equality, diversity and inclusion (EDI) have come into stark focus. In the emergency, many families were unable to access or use the necessary technology, thus widening the 'digital divide' and raising questions of fairness and social justice. Unless these wider societal concerns are addressed, there is a risk that the potential for EdTech and EdAI to bring about much needed innovations in multilingual education will be undermined.

In setting the scene, I ask in what ways the field of multilingual learning, teaching and assessment can be changed for the better, and I flag up some specific challenges and opportunities that are already emerging in a post-Covid world.

Nick Saville is a founder member of the Association of Language Testers in Europe (ALTE) and is the elected Secretary-General. He is Director of Research & Thought Leadership at Cambridge English (University of Cambridge) and sits on several University Boards, including: the Interdisciplinary Research Centre for Language Sciences; the Institute for Automated Language Teaching and Assessment; and English Language iTutoring (ELiT), providing AI-informed automated systems. He was a member of the Advisory Council for the Institute for Ethical AI in Education whose final report was published in March 2021.

He is a consultant for European institutions including the Council of Europe, the European Parliament & the EU Commission. In this capacity, he has been involved in the Common European Framework of Reference for Languages (CEFR) since its inception and has led several projects in the field of multilingualism. He was a founding associate editor of the journal Language Assessment Quarterly and is currently joint Series Editor of the Studies in Language Testing series (Cambridge Assessment English/Cambridge University Press) with Professor Lynda Taylor.

1ST INTERNATIONAL DIGITAL SYMPOSIUM
#ALTEdigital2021

Innovations and implications of language assessment technologies

Paula Buttery, University of Cambridge

Assessment is an intrinsic aspect of the language learning experience: whether it be assessment in the form of routine monitoring by a teacher in the classroom; assessment in the form of regular progress and placement testing; or the certified assessment of high-stakes examinations. Today, in the Covid era, the motivation for automating some aspects of this assessment has never been so well understood. Language assessment technologies are acutely needed to facilitate the remote classroom, provide monitoring feedback and even to deliver examinations.
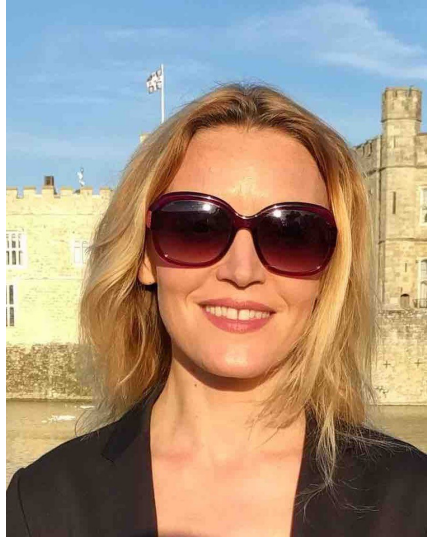
Technologies have been in development for some years that emulate components of assessments that are repetitive and not a good use of an expert's skills or time. These artificial intelligence technologies have been deployed to remove the mundane aspects of assessment, freeing up teachers and examiners to carry out more specialised and interesting work. A positive side-effect of using such technology is that statistical information about individual students and student cohorts is easy to collate: if presented correctly, this can provide useful insights into learning journeys and inform strategies for maximally effective teaching.

Pre-Covid, the crucial expert personalised ingredient that teachers and examiners bring to assessment was taken as given. A question then is whether technological advances can support teachers and examiners in an entirely remote setting.

In this talk I will present the work of the Cambridge Institute for Automated Language Teaching and Assessment (ALTA), who have been working on the automatic assessment of English language learners from classroom monitoring to high-stakes examination. I will discuss the key technologies involved in this work; it's transferability to other languages; explain what aspects of language assessment are easy/hard for artificial intelligence; and discuss the ethical issues in creating digital representations and judgements of students.

Paula Buttery is a Professor of Language and Machine Learning at the Department of Computer Science and Technology, University of Cambridge. She has expertise in developing computer applications (natural language processing) and also in language cognition (computational psycholinguistics). She is currently the Director of the Cambridge Institute for Automated Language Teaching and Assessment (ALTA). This is an artificial intelligence institute that uses techniques from machine learning and natural language processing to improve the experience of language learning online. Her work within ALTA focuses on English language learners but she also has experience working with under-resourced and endangered languages.

Multilingual assessment integrated with language policy: Lessons from research and new initiatives for equality and justice

Elana Shohamy, Tel Aviv University

Students enrolled in schools where the language of instruction is different than the language they acquired proficiency in from their local communities or country of origin face major difficulties in their school achievements. Research reveals that it takes up to 10 years for these students to reach similar academic achievements as that of native speakers of the dominant language. Indeed, indigenous and immigrant students continue to rely on their home languages as important resources when they are given this opportunity; studies show that the presence of their home languages along with the new language on tests enhances their academic achievement significantly (Shohamy 2011). This testing method represents an additive approach versus a subtractive one where home languages of immigrants and indigenous students are absent in schools.

This paper reports on a large-scale research programme, the aim of which was to develop a new multilingual policy in Israeli schools (Shohamy and Tannebaum, 2020). In the testing component of the research, bi-multilingual test batteries in reading and writing in science were developed to examine their validity and practicality. This study is innovative in that it uses a variety of methods for data collection – think-aloud protocols, focus groups, attitude questionnaires and interviews. The study also examined differences and similarities of academic performances of two different groups of L2 users, immigrant and indigenous groups. The results enriched our knowledge about the suitability of bilingual tests for multilingual educational policies and realities. Furthermore, they demonstrated how these testing methods introduce justice and inclusion as well as maintenance of home languages and translanguaging competence.

Elana Shohamy is a professor of Language Education at Tel Aviv University School of Education where she teaches and researches issues of justice and rights as they relate to language testing, language policy, immigrants and minority groups as well a linguistic landscape; her current testing research focuses on multilingual assessment as part of a new multilingual educational policy in Israel. She authored The power of tests: A critical perspective on the uses of language tests (Routledge, 2001), Language policy: Hidden agendas and new approaches (Routledge, 2006), and edited the Testing and Assessment volumes of The Encyclopedia of Language and Education (Springer, 2009 and 2018). She was granted the ILTA Lifetime Achievement Award in language testing (2010) for her work on critical language testing.

Can tests be fair and valid?

Barry O'Sullivan, British Council

There are a number of significant issues that undermine much of what we do in language testing. These issues spring from the fact that test developers have, for many decades, failed to come to terms with the critical concept of construct definition. The most obvious failure is that of developers who employ measurement-based approaches that can never work. This issue has led to the development of construct-agnostic tests that continue to dominate some traditions. Another significant issue is the decontextualization of the definition of the language construct. Here, language is seen as a cognitive function, with little or no recognition of the social factors that change language from a knowledge-focused ability to its use-focused reality. The tests that result from such a limited definition are similarly problematic to measurement-based tests in that neither can be shown to be either fair or valid.

In presenting the above, I will argue that construct definition lies at the heart of all of the issues identified by the organisers of this conference as being central to the concepts of fairness and validity. This is because the language (or other) ability we are testing lies within the candidate (it is not simply an abstract construct) and that the test taker is part of a broader social and educational context. Therefore, any definition (and by inference any claim that it is fair and valid) must take these broader contexts into consideration. Any test must fit conceptually within the learning system for which it is intended while at the same time reflecting sound development and validation practice. I conclude that the sociocognitive approach offers such a conceptual platform and at the same time represents the basis of broader concepts such as localisation and personalisation.

Barry O'Sullivan is the Head of Assessment Research & Development at the British Council. He has undertaken research across many areas on language testing and assessment and its history and has worked on the development and refinement of the sociocognitive model of test development and validation since 2000. He is particularly interested in the communication of test validation and in test localisation. He has presented his work at many conferences around the world, while almost 100 of his publications have appeared in a range of international journals, books and technical reports. He has worked on many test development and validation projects over the past 25 years and advises ministries and institutions on assessment policy and practice.

He is the founding president of the UK Association of Language Testing and Assessment (UKALTA) and holds honorary and visiting chairs at a number of universities globally. In 2016 he was awarded fellowship of the Academy of Social Science in the UK and was elected to Fellowship of the Asian Association for Language Assessment in 2017.

Language testing ethics, and our place in the ecosystem

Bart Deygers, Ghent University

In this talk we will explore the different roles of language testers in a larger societal-political ecosystem. Delineating and defining these roles is a sobering exercise, but one that helps us understand our role as ethical actors in society and within our field. One that helps us see why our ethical codes have had limited actual impact on our practice.

This talk will be an exercise in demystification. We will consider the connection between human rights and language testing codes, discuss the hypocrisy of power, and the appeal of non-binding agreements with limited enforceability. But we will end on a positive note: what can we do to have an actual positive impact on the ecosystem in which language testing exists?

Bart Deygers is Professor of Second Language Didactics and Assessment at Ghent University. His primary research interests include language testing policy, equal educational opportunities for vulnerable learners, and language testing for migration purposes. In his research he examines and questions the role and impact of high-stakes tests in today's society. Human rights, justice and fairness are central to his publications and research, which focus on empirically verifying the policy makers' assumptions regarding test score use and on examining how language testing policies take shape.

OECD: The PISA 2025 Foreign Language Assessment

Catalina Covacevich, OECD

The OECD's Programme for International Student Assessment (PISA) is a triennial international survey, which aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students. The tests assess to what extent students can apply their knowledge to real-life situations and be equipped for full participation in society. Background questionnaires provide context that can help interpret the results.

The PISA 2025 cycle will include for the first time an optional assessment of foreign languages. The assessment will provide policy makers and educators with comparable results of their students' foreign language competence and allow them to gain insights into the best practices and policies for teaching and learning a foreign language. While the first cycle will focus only on English, further cycles are expected to incorporate more languages.

I will provide a brief description of PISA and then discuss the design of the PISA 2025 Foreign Language Assessment, mentioning how it relates to safeguarding the future of multilingual assessment in the COVID world.

Catalina Covacevich joined the OECD in 2016 as a policy analyst to support the implementation of the PISA for Development initiative. She is currently working on the development of the PISA 2025 optional Foreign Language Assessment as well as supporting the design and implementation of PISA. Before joining the OECD she worked for the Inter-American Development Bank as a Local Education Specialist in Chile, for the Assessment Unit of the Chilean Ministry of Education, and for MIDE UC, a Chilean institute specializing in large-scale educational assessments, where she participated in the design and implementation of national and regional assessments. She holds a MSc in Social Psychology from the London School of Economics and Political Science and speaks English, Spanish, and some French.

Eaquals: Looking back, looking forward – Lessons learned from the ongoing Covid emergency

Lou McLaughlin, Eaquals

Eaquals is the leading professional association for language education throughout the world. It is an international non-profit membership association. Founded in 1991, we are independent of any commercial group and exist solely to serve the interests of language learners and the language education profession. Eaquals members – organisations and institutions involved in the teaching and learning of all languages – pursue the highest standards. Those of Accredited Members are regularly verified by means of rigorous and thorough inspections. Among its many publications, Eaquals publishes its full inspection scheme (version 7.2) and those extensions which cover Higher Education. It also provides language teaching resources and worldwide training and consultancy services. Eaquals also has Associate Members (institutions such as exam boards, publishers, national associations, quality assurance organisations, and so on), individual members, project members, and partnerships.

Eaquals has built on its many MoUs with international organisations. ALTE is one of the longest standing partners of Eaquals. Together we developed the very first European Language Portfolio for adults, and the first electronic version.

Currently, we are working jointly to create online resources to promote and certificate assessment literacy, at various levels of competence that are closely integrated with the European Profiling Grid.

Eaquals proposed establishing a survey to the ECML in Graz on how language educators are coping with and learning from the challenges of Covid, to be followed up with think tanks. The ECML, in co-operation with its Professional Network Forum, launched a Europe-wide project seeking to gather information and exchange views for the survey and think tanks.

This survey took place in February. The results will be used in a series of online think tanks, a Colloquium and

a publication. This process is supported by the ECML Secretariat. Our initiative has the backing of the European Commission which has generously contributed funds to support it. A report on the think tanks and this survey will be published in late 2021 or early 2022.

Lou McLaughlin is the Executive Director of Eaquals (Evaluation & Accreditation of Quality Language Services). She is the Founder Chair of ELT Ireland, the network for ELT professionals in Ireland, and is on the IATEFL Board of Trustees as the Associates Representative as well as serving as a Trustee for the Bell Education Trust. She holds a PhD in Applied Linguistics, Masters in ELT, DELTA, TESOL Cert. in Leadership & Management and Trinity TESOL Cert. Her special interests are Young Learner teacher training, teacher cognition and management in ELT. She wrote and developed the online YL Management Training course run by International House (IH) London and worked as an inspector for International House World Organisation (IHWO). She is a frequent speaker at international conferences and was the co-editor of Children Learning English: From Research to Practice (2015).

Council of Europe Education Policy Division: Plurilingual and intercultural education for democracy

Michael Remmert and Ahmet-Murat Kılıç, Council of Europe Education Policy Division

This talk will outline the aims and principles of the Council of Europe's language policy and inform participants about recent developments, in particular the publication of the CEFR Companion Volume. The protection and promotion of Europe's rich linguistic and cultural heritage have been a cornerstone of the Council of Europe's work since the entry into force of the European Cultural Convention in 1954, designed to foster among the nationals of all state parties the study of the languages, history, and civilisation of the others, and of the civilisation which is common to them all. Access to this heritage is greatly facilitated through the provision of lifelong quality language education which is based on learners' needs and is inclusive, plurilingual and intercultural. Moreover, quality language education is fundamental to the effective enjoyment of the right to education and contributes to democratic citizenship and social cohesion. The CEFR, in turn, has played a key role in promoting quality plurilingual education, facilitating greater social mobility and stimulating reflection and exchange between language professionals for curriculum development and in teacher education. How-

ever, its application has arguably in some contexts led to consequences not intended by the CEFR, such as creating barriers to the integration of migrants. In this regard, there is a case for reconsidering the purpose and modes/ways of language assessment to meet the principle of fairness and prevent inequalities and violations of individual rights, as well as to contribute to effective language education (assessment as learning and for learning). Through the publication of the CEFR Companion Volume, the Council of Europe intends to give equal importance to learning and teaching and to introduce alternative ways of assessment such as the language proficiency profiles, which are seen as a way of properly assessing individual language proficiency across four modes of communication and in various languages.

Ahmet Murat Kılıç works as a Programme Manager at the Education Policy Division of the Council of Europe. He manages three intergovernmental programmes, i.e. Language Policy, including the promotion of plurilingual education, the publication, promotion and dissemination of the CEFR Companion Volume, and contributing to the educational inclusion of Romani children; Digital Citizenship Education (DCE), including co-ordinating the implementation of the Committee of Ministers Recommendation on developing and promoting digital citizenship education through the DCE promoters' network; and Artificial Intelligence in Education. He joined the Council of Europe in 2011, initially working on Education for Democratic Citizenship and Human Rights Education (EDC/HRE) in the Ankara Programme Office. He managed the European Union and Council of Europe Joint Project on EDC/HRE in Turkey. In 2017 he moved to the headquarters and took charge of the language policy programme. He holds an MBA from the University of Nevada, Reno and specialises in project management. He previously worked in the private sector in various fields including energy, automotive, maritime and import/export.

Michael Remmert is the Head of the Education Policy Division of the Council of Europe (CoE). He manages and supervises the three pillars of the intergovernmental Education programme: Competences for Life in Democracy, including Digital Citizenship Education; Inclusive Approaches to Education, including Language Education Policy; and Ethics and Integrity in Education, including higher education. He oversees the activities of the Steering Committee for Education Policy and Practice (CDPPE) and other committees and working groups for the development and implementation of the education programme. He also manages and supervises the CoE's Network of Schools of Political Studies (SPS), which delivers high-quality training for young political leaders in member states and countries in the CoE's neighbourhood.

From 2011 to 2018, he was Deputy Head of the Directorate of Policy Planning (DPP). This Directorate's mandate included addressing emerging opportunities and challenges for Europe with interlocutors internal and external to the CoE, thereby fostering forward-looking and transversal thinking and action.

A L T E Association of Language Testers in Europe.

Holding a doctorate in Political Science and Economics from the University of Freiburg/Germany, he joined the Council of Europe in 1994, initially working in the field of social cohesion, then on democratic institutions and e-governance.

Council of Europe: European Centre for Modern Languages (ECML)
The dynamics of multilingual assessment

Sarah Breslin and Jose Noijons, European Centre of Modern Languages

In its 25 years of existence the European Centre of Modern Languages (ECML) has focused on various aspects of multilingual assessment. Many of its projects have aimed at assisting teachers and testers in various aspects of assessment, such as the assessment of reading and writing in the primary language classroom on the basis of CEFR descriptors. Other projects have provided support to language professionals to become familiar with the philosophy of learning, teaching and assessment, in line with the CEFR, or have guided teachers in the planning, implementation and evaluation of whole-school projects. Another of the ECML's initiatives, in cooperation with the European Commission, is RELANG which assists European countries not only in aligning tests and examinations to the CEFR and in creating CEFR-based examinations, but also in aligning existing curricula to the CEFR and in creating new CEFR-based curricula. Its most recent activities are in the area of exploring ways to test mediation and plurilingual skills in the classroom.

Since October 2013, Sarah Breslin has been the Executive Director of the

European Centre for Modern Languages (ECML), an institution of the Council of Europe, based in Graz, Austria. A passionate linguist with a thorough understanding of both policy and practice in language education and general education, Sarah has worked in a range of sectors and countries since she graduated with first class Honours in French and German from the University of Glasgow in 1986. After training to become a language teacher, she worked for 15 years in Catalonia as an EFL teacher and became first Director of Studies and then Director of a prestigious language school in Tarragona. She then moved into Higher Education, where she was Head of Modern Languages in the Faculty of Tourism and teacher of English and German. On returning to the UK, she worked for four years in further education, teaching English for Speakers of Other Languages (ESOL) and Spanish, before becoming Head of International and EU programmes. Before taking up the post as Director at Scotland's National Centre for Languages at the University of Strathclyde (SCILT), she worked for the Scottish Qualifications Authority (SQA) where her focus was on education policy both at UK and EU level, as well as cross-border qualification recognition. She recently achieved her Doctorate degree focusing on the contribution of the ECML to the professional growth of language teacher educators.

José Noijons has developed a range of skills in the sampling of student performance, data collection, data analysis and reporting in the context of programme evaluation, both on a national level and in an international context, in national assessment programmes in a number of countries, and in an international context, such as PISA. His primary training has been in language testing. From there he has widened his scope and has dealt with tests of mathematics, (social) sciences and other school subjects. He has been a staff member at Cito, the Dutch National Institute for Educational Measurement, for 33 years. After his retirement he continued his work for the European Centre for Modern Languages (ECML) in Graz, Austria. The ECML is a Council of Europe institution. At present he coordinates a joint EU-Council of Europe project: RELANG, assisting educational authorities in relating language examinations and curricula to the levels of proficiency defined in the CEFR. In this capacity he has worked in a number of European countries: Albania, Armenia, Cyprus, Estonia, Greece, Hungary, Iceland, Latvia, Lithuania, Malta, Montenegro, Romania, Spain and Slovenia. He has coordinated the RELANG activities in other European countries, such as Belgium, Croatia, the Czech Republic, Denmark, FYROM, Ireland, Italy, Norway and Portugal.

The use and impact of the ALTE Principles of Good Practice

► Waldemar Martyniuk and invited guests

In April 2020, ALTE launched its revised Principles of Good Practice. The document is freely available to download from the ALTE Resources webpage: www.alte.org/Materials. This is the third edition of a document that was first put together in 1994. The document is intended as a coherent set of guidelines that all ALTE Members and Affiliates subscribe to when creating and running language tests. The panel on The Use and Impact of the ALTE Principles of Good Practice is organised to offer insights into ways in which the guidelines contribute to the improvement of quality and enhancement of professionalism in member organisations/institutions. Representatives of ALTE member organisations as well as partner associations contribute to the panel by offering brief critical reflections and comments on the ALTE Principles of Good Practice 2020.

Contributions:

- Introduction: The ALTE Principles of Good Practice 2020 – Origins - Principles - Practice - Support – Nick Saville & Waldek Martyniuk

- The ALTE Principles of Good Practice in use: accounts by Dina Vilcu, Babeş-Bolyai University, Cluj-Napoca, Romania, and Kateřina Vodičková, Charles University, Prague, Czech Republic

- The ALTE Principles of Good Practice 2020: critical reflections and comments by Salomé Villa Larenas (representing LAALTA) and Peter Lenz (representing EALTA)

Waldemar Martyniuk is Professor and Executive Director at the Institute of Polish Language and Culture for Foreigners of the Jagiellonian University in Krakow, Poland. He holds a PhD in Applied Linguistics and is a teacher trainer, author of several textbooks, curricula, and testing materials for Polish as a foreign language. He is a Visiting Professor and lecturer at universities in Germany (Bochum, Giessen, Goettingen, Mainz, Muenster), Switzerland (Basel), and in the USA (Stanford University). He was seconded to the Council of Europe, Language Policy Division in Strasbourg, France, 2005–2006, and 2008–2013 as Executive Director of the European Centre for Modern Languages in Graz, Austria. Since 2019, he has been Chair of the Board of Trustees at ALTE.

Is large-scale testing antithetical to learning-oriented language assessment?

► Anthony Green, Centre for Research in English Language Learning and Assessment (CRELLA), University of Bedfordshire

► Jessica R W Wu, Language Training & Testing Center, Taiwan

► Liying Cheng, Queen's University, Canada

► Nick Saville, Cambridge Assessment English

► Atta Gebril, American University in Cairo

With its emphasis on informing and promoting learning, learning-oriented assessment (LOA) is usually identified with classroom practices and is sometimes contrasted with external testing for accountability purposes (Carless, Joughin and Mok 2006). This panel meeting turns the spotlight on large-scale proficiency testing, considering the extent to which large-scale testing currently supports effective language learning, how testing might be reformed to better support learning and how new technologies might contribute.

The panel brings together perspectives from testing agencies, institutional testing programmes and classroom research to address the question in relation to the three pillars of LOA: learning tasks, learner agency and feedback.

Learning tasks engage learners in an interactive process, providing opportunities for learners to reconsider and revise their responses.

Learner agency implies that learners are not passive participants in assessment, but exercise choice and a degree of control over the content and direction of learning. This embraces practices such as self- and peer-assessment.

Feedback not only involves the provision of information to learners about their performance, but also the use of such information to inform and enhance teaching and learning. The panellists will report on and debate the steps that test developers can take to encourage positive washback through task design, support for educators and enhanced score reporting. They will consider the impact of such measures on learners and other

stakeholders in national education systems and reflect on new directions for assessment in technological environments.

Anthony Green is Director of the Centre for Research in English Language Learning and Assessment (CRELLA) and Professor in Language Assessment at the University of Bedfordshire, UK. He has consulted and published widely on language assessment issues. His main research interests lie in the relationship between assessment, learning and teaching and in learning-oriented assessment.

Jessica R W Wu holds a PhD in Language Testing. She is the R&D Program Director at the Language Training & Testing Center (LTTC), a non-profit educational foundation in Taiwan. She also serves as an adviser to the government on the development and administration of L1 tests. She has published numerous articles and book chapters in the field of language testing and has presented her work at conferences around the world. She is currently the President of the Asian Association for Language Assessment (AALA, 2018–2019).

Liying Cheng is Professor and Director of the Assessment and Evaluation Group (AEG) at the Faculty of Education, Queen's University, Canada. Her primary research interests are the impact of large-scale testing on instruction, the relationships between assessment and instruction, and the academic and professional acculturation of international and new immigrant students, workers, and professionals. Her seminal research on washback focuses on the global impact of large-scale testing.

Nick Saville is a founder member of the Association of Language Testers in Europe (ALTE) and is the elected Secretary-General. He is Director of Research & Thought Leadership at Cambridge English (University of Cambridge) and sits on several University Boards, including: the Interdisciplinary Research Centre for Language Sciences; the Institute for Automated Language Teaching and Assessment; and English Language iTutoring (ELiT), providing AI-informed automated systems. He was a member of the Advisory Council for the Institute for Ethical AI in Education whose final report was published in March 2021.

Atta Gebril is Associate Professor and MATESOL Program Director at the Department of Applied Linguistics, American University in Cairo (AUC). He also serves as chair of the Institutional Review Board at AUC. He obtained his PhD in foreign language and ESL education with a minor in language testing from the University of Iowa. He previously worked for American College Testing (ACT, Inc.) from 2005 to 2006 where he was part of the Workkeys team. In addition, he has been working as an assessment and evaluation consultant and has participated in several test development projects in many parts of the world. His research interests include writing assessment, reading–writing connections, assessment literacy, and test validation. His dissertation work focused on score generalizability of academic writing tasks. He serves as an associate editor of Language Assessment Quarterly and he is also on the editorial boards of Assessing Writing and the TESOL Journal. He also recently edited a volume on learning-oriented language assessment, which was published by Routledge in 2021. In 2019, he won the 'Best Article in the Field of Language Testing' award from the International Language Testing Association.

Assessing less commonly taught languages in times of declining enrolments and Covid: Addressing challenges and creating opportunities

► Koen van Gorp, Center for Language Teaching Advancement

► Dina Vilcu, Babeș-Bolyai University, Romania

► Ina Ferbežar, Centre for Slovene as a Second and Foreign Language, Slovenia

► Emyr Davies, CBAC-WJEC, Wales

Enrolments in language courses at institutions of higher education seem to have been steadily declining in Europe and the USA over the past two decades, leading to what has been referred to as a language crisis. However, in times of Covid, more and more people seem to have taken up learning another language. As language teaching and learning have turned almost fully online, administering language assessments, and certifying students in a Covid-safe environment has proven to be a challenge, especially for many less commonly taught languages (LCTLs). Whereas large international tests of more commonly taught languages often had the resources to move tests quickly online and provide web-based proctoring environments worldwide, test providers of LCTLs had the advantage of being able to adapt more quickly and respond to changing demands of stakeholders and the possibilities offered by technology. Additionally, the pandemic seems to have created a situation that encourages teachers and testers of LCTLs to reconsider the traditional role and ways of assessment and think of alternatives.

This panel aims to explore how language testers of LCTLs have responded to the challenges and opportunities for language assessment during the pandemic, and how they might help safeguard the future of multilingual language assessment in the (post) Covid world. The panel will consist of experts in the assessment of LCTLs in Europe and the USA. The presenters will reflect on the challenges, and opportunities or solutions in classroom-based assessment of different LCTLs (Van Gorp) and the development and administration of standardized tests of Welsh, Romanian, and Slovenian (Davies, Vîlcu, and Ferbežar, respectively). Speaking to

four different contexts, the presenters will discuss common solutions that could be adopted by teachers and testers of LCTLs facing similar challenges.

Koen Van Gorp is Head of Foreign Language Assessment at the Center for Language Teaching Advancement, and core faculty in the Second Language Studies Ph.D. Program at Michigan State University. Koen serves on the TOEFL Committee of Examiners (2020–2024) at Educational Testing Service (ETS) and is an Individual Expert Member of ALTE. From 2010 until 2015, he was the Director of the Certificate of Dutch as a Foreign Language at KU Leuven (Belgium). He is co-Editor of TASK. Journal on Task-Based Language Teaching and Learning. His research interests are task-based language teaching and assessment, and multilingual education and assessment.

Dina Vîlcu is a senior lecturer at Babeș-Bolyai University, Cluj-Napoca, Romania. Her main activities and responsibilities are related to teaching, assessing, and research in the field of Romanian as a foreign language (RFL). She coordinated the assessment segment in the 3-year project Romanian as a second language (RLNM), and she was in charge of the auditing process of the RFL examinations (Levels A1, A2, B1 and B2) by the Association of Language Testers in Europe (ALTE). She organised assessment related international conferences (ALTE 51st Meeting and Conference Day, 11–13 April 2018). She has coordinated the Language for Specific Purposes Special Interest Group (LSP SIG) of ALTE since 2016. She has worked on a number of volumes dedicated to assessing

RFL or language for specific purposes.

Ina Ferbežar is a head of the Examination Centre at the Centre for Slovene as a Second and Foreign Language (Faculty of Arts, University of Ljubljana, Slovenia). She was executive head of the Centre for Slovene (2008–2012). She is active in the field of language policy (contributor to the White Book on Education (2011) and National Program for Language Policy (2011), co-editor of the Slovenian language policy portal). She was a member of the editing group and editor of the Slovenian version of the CEFR, and chaired the committee to prepare tests for Slovenian as a L1 within the national examination programme (2013–2017). She is a representative of the University of Ljubljana in ALTE (a member of the Executive Committee, 2009–2016, and Standing Committee, 2015 to present). Her research explores language testing and language policy, text comprehension and text comprehensibility.

Emyr Davies is responsible for a suite of exams for adult learners of Welsh at CBAC-WJEC, an award organisation. He graduated in Welsh, has a higher diploma in Linguistics and a PhD. He worked as a lecturer in a college in west Wales, teaching language and literature courses, before starting his present post. He has represented CBAC-WJEC in ALTE since 2001, in respect of the Welsh for adults exams. He has contributed to several ALTE conferences and meetings and is a member of the Board of Trustees. His current research involves predicting item-difficulty in the assessment of Welsh, adapting speaking tests for the online environment, and investigating the possibilities of remote invigilation.

What's the point of assessment? Making it fit for humans

Amy Devine and Tom Booth, Cambridge Assessment English

The Covid-19 pandemic has disrupted educational systems across the globe, which has required a rapid shift to online teaching and assessment in many contexts. This shift has accelerated reflection on current assessment practices and the need to innovate in order to meet the changing needs of test users.

Nowadays, of course, the scope of what we call 'assessment' stretches far beyond the image of rows of exam candidates taking an exam. Rather than the 'summative' and 'formative' distinction, the principles and practices of assessment can be seen as serving many different purposes in supporting and promoting learning (Jones and Saville 2016).

At Cambridge Assessment English, we explore the possibilities afforded by different forms of assessment meeting different needs and serving different purposes. Many of the products and exploratory projects we work on are digital and focus on integrating learning and assessment. Whereas the temptation might be to start from tried-and-tested types of content and assessment tasks, we are now thinking increasingly differently, retaining our principles of learning-oriented assessment but applying them in more innovative ways to meet the needs of our users.

In this session, we will explore how games-based assessment, gamification and innovative task design can be used in the development of assessments for different purposes. It will be a highly interactive session in which we will share some of our experiences, and participants will be invited to draw on their experience and expertise as well as think in new and creative ways. This will be an opportunity for participants to apply assessment principles to new contexts and opportunities, and perhaps to challenge their usual ways of thinking about assessment.

Amy Devine has been working as a Senior Research Manager at Cambridge Assessment English since 2016. She manages research on assessment and learning products including games-based assessments, digital products and examinations for young learners. Her research interests include the role of affective variables in learning, with a particular focus on test-taker emotions, motivation and engagement. Prior to joining Cambridge Assessment English, Amy worked on several funded research projects in education and psychology at the University of Cambridge. She completed a PhD in Psychology at the University of Cambridge in 2018.

Tom Booth has been working on digital learning and assessment products at Cambridge Assessment English since 2010. He currently works on integrated learning and assessment in a team that includes software developers, designers and assessment experts, and is particularly interested in how digital assessment and detailed feedback can be embedded into learning products in engaging and impactful ways. Before joining Cambridge Assessment English, he taught English as a foreign language to teenagers and adults.

From watching to reaction and finally to action: A brief history of remote proctoring

Vincent Folny, France Education International

In 2020, the pandemic undoubtedly modified our agendas. It made us understand what was resistant, appropriate, obsolete, or missing in our exams. After a year, it seems interesting to us to analyse the situation concerning proctoring before the crisis, during the crisis and what could be the path followed for the next years. During this presentation, we will examine the procedures and tools that were used before the crisis, what was put in place during the crisis and what we plan to do in the future. Through a brief analysis of risks and opportunities, we will analyse what could explain the decisions of France Education International. We will see that the state of development, the availability and the appropriation of the technology, and the legal and ethical considerations were crucial in explaining our decisions. We will try to analyse why certification bodies should not confuse all technology development with reaction, or reaction with medium- and long-term action. A rational management of these different temporalities is probably necessary to avoid technological ignorance, lack of proportionality with the pandemic reactions or bad management of medium- or long-term strategies.

After studies in France and Canada focused on French as foreign language, assessment and psychometrics, **Vincent Folny** has been working at France Education international (FEI) for 14 years. FEI is the public operator for the Ministry of National Education which is dedicated to the certification of the French language. He oversees and monitors test quality and test development (from content, constructs and measure-

ment perspectives). He has attended ALTE meetings for more than 13 years and has been chair of the Quality Management System (QMS) and the Special Interest Group (SIG) on technology. During the last decade, he acted as an ALTE auditor. Since 2021, he is the chair of the Standing Committee.

Personalised assessment

Mark Elliott, Cambridge Assessment English

The concept of personalised assessment has an established history. Computer Adaptive Testing (CAT), in which new items are selected on the basis of performance on previous items, has been around since the 1980s, and proved its worth as a means of providing efficient tests tailored to individuals. Modern computational techniques now allow for this approach to be broadened in a number of ways.

I will introduce this workshop by reviewing the history and approaches of CAT and how similar approaches might apply to multidimensional cognitive models. I will then draw a sharp distinction between standardised tests and online adaptive learning for formative learners, and how the latter context permits greater scope for personalisation, focusing in particular on how this might apply to personalisation of the interpretation of assessment results as well as to the item selection process. Finally, I will consider a more radical scope of personalised assessment situated within a wider conceptualisation of assessment which encompasses other forms of evidence-gathering beyond test scores, focusing on the use of corpus linguistics techniques to profile written language to create a benchmarked profile of learner language to complement constructed measures of ability. There will be time for interactive discussion and sharing of experiences from the audience.

Mark Elliott is a Senior Assessment Quality and Validity Manager at Cambridge Assessment English, where he primarily works on test design, both in terms of test construct and psychometric properties, and Integrated Learning and Assessment, developing approaches and models to personalised adaptivity and diagnostic models. Within Cambridge Assessment English, he has previously worked in the Assessment division, the production of speaking tests and modified materials for candidates with special requirements, and in Research and Validation, where he was responsible for research into listening and contributed several chapters to volumes in the Studies in Language Testing series. He is also a member of the Automated Language Teaching and Assessment Institute (ALTA), where he is a PhD candidate within the Department of Computer Science and Technology at the University of Cambridge, where his research focuses on adaptive learning systems and accounting for rater variability within examiner-marked essays.

Using data

Andrew Caines, University of Cambridge

I will introduce this workshop with an overview of good practices in data retention, management and analysis for education applications. I will talk generally about educational settings and the importance of good data practices, with a particular focus on language testing scenarios and learning-oriented assessment. I will describe how data analysis can impact language testing services, and give some case studies of data science in practice. For instance, what can we learn about tests and test-takers from data science analyses? How can we use test data to personalise further learning activities and formative feedback for test-takers? What kind of data should we collect in order to make generalisations and continue improving testing techniques? I will underline the link between careful data curation and the successful development of AI systems, and consider the ethical implications therein. We will have time for interactive discussion and sharing of experiences from the audience.

Andrew Caines is a Senior Research Associate in the Department of Computer Science and Technology and the Automated Language Teaching and Assessment (ALTA) Institute at the University of Cambridge, UK. He has been with the ALTA Institute since its foundation in 2013, and has published on spoken learner English, natural language processing and machine learning for assessment and teaching purposes, and the development of education technology. He also works on projects relating to spoken English grammar, the automatic detection of grammatical errors in written English, and the listening construct in cognitive terms. He obtained his PhD on innovation in spoken English from the University of Cambridge in 2010 and has had an active publication record since then.

What is reading and writing in the digital world? Redefining constructs for special needs

Hanne Lauvik and Stine Sørlie, Skills Norway

In 2020, Skills Norway was asked by the Ministry of Education and Research to conduct trials of assistive technology tools for dyslexic candidates in the test of Norwe-

gian for adult immigrants, and to consider amending the regulations for special arrangements to include these tools. This was on the initiative of adult education centres and organisations working to improve conditions for those struggling with dyslexia, who had voiced concerns that dyslexic test takers would not be able to show their full language potential in the reading and writing parts of the test.

Digital tests have the possibility of using a range of different assistive technology tools to support reading and writing. Spell check, grammar check, thesaurus and dictionaries, word prediction, next word prediction, text-to-speech and speech-to-text programs, are some examples of tools commonly used in study and work settings today, and not exclusively by those struggling with dyslexia. For language testers, decisions about which tools to permit, if any, are challenging, and will often raise questions about what reading and writing skills in today's digital world are, what we can and should test, and how to do it.

Skills Norway sought to base the decisions about changing the current test regulations not just by evaluating the information gathered from the trials with assistive technology tools, but also by reviewing and questioning existing test constructs and regulations for special needs with regards to validity, fairness and justice. In this workshop we will focus on group discussions about questions that arose in this process of amending special arrangements for the reading and writing test. The workshop will precede a presentation later on in the symposium, where we will present more information about the trials conducted, and the new arrangements offered to dyslexic test takers.

Stine Sørlie and **Hanne Lauvik** work at Skills Norway, a directory belonging to the Norwegian Ministry of Education. Skills Norway is responsible for developing the test in Norwegian for adult immigrants, a digital test that measures at Levels A1–C1. They work together to ensure that test takers with special needs have an equal opportunity to document their language skills with good quality tests. An important part of this work is to develop different test alternatives, both digital and pen-and-paper tests, to address different needs, as well as making sure that the test regulations for special needs are fair and just. They are also responsible for the tests offered to immigrant sign language users, as well as planning and running annual seminars for sign language teachers and raters on assessment of sign language skills according to the CEFR. A recent focus has been to revise and amend the arrangements for dyslexic candidates in digital reading and writing tests.

GDPR and remote proctoring

Ian Hanahoe, Cambridge Assessment

During the Covid-19 pandemic the way students are assessed has changed dramatically and awarding bodies have had to change their processes to accommodate the needs of students. One way of addressing these needs has been the use of remote proctoring, which has inherent risks when it comes to the processing of students' personal information. In order to comply with data protection legislation, it is essential awarding bodies carefully consider the implications of remote proctoring through robust due diligence. This will, more often than not, involve the use of Data Protection Impact Assessments where it will be necessary to consider aspects such as the legal basis for the processing, the use of biometric information and the implications of using human or AI proctoring. This session is designed to give delegates an understanding of these issues.

Ian Hanahoe joined Cambridge Assessment in November 2019 as the Data Protection Officer. He has overall responsibility for ensuring Cambridge Assessment is in compliance with data protection legislation around the world through the use of Data Protection Impact Assessments, training of staff, dealing with Subject Access Requests, and providing advice and guidance to colleagues. Before joining Cambridge Assessment, he spent 18 years at the University of Hertfordshire (UH). He supported UH in a number of different legal areas including copyright, intellectual property, corporate law and commercial law he also acted as the Legal Advisor for one of the university's subsidiaries. In 2011 Ian was appointed the Information Compliance Manager for the university dealing with data protection and freedom of information requests. In the lead up to the introduction of the GDPR in 2018, Ian spearheaded the university's preparation for the change in legislation to ensure compliance was maintained, and in March 2018 was appointed the Group Data Protection Officer.

Implications of the CEFR-CV for language assessment

Brian North, Eurocentres Foundation

After briefly reminding participants of updates to the CEFR 2001 descriptors, this talk will outline various implications of the CEFR for language assessment, largely overlooked in 2001 and highlighted more clearly in 2020.

This starts with the CEFR model of language use itself, which is complex and integrative rather than linear and

Cartesian, emphasising the combination of reception, production, interaction and mediation in natural language use and the agency of the user/learner as a collaborative social being ('social agent'). This model presents challenges to language assessment. In fact, after a brief flirtation with both partial competences and integrated skills at the start of the communicative approach, language assessment prefers for operational reasons to focus on single skills and discrete tasks for isolated individual candidates.

It is easier to see how collaborative, integrated, action-oriented tasks can be used for classroom formative/continuous assessment with the aid of descriptors (assessment as learning; assessment for learning) than it is to solve the complications caused by assessing individuals in formal examinations (assessment of learning) on the basis of performance in collaborative tasks, where they are constrained at least to some extent by the behaviour and performance of others. This conundrum is perhaps one reason why mediation, where it has been operationalised in assessment (in Germany and Greece), has tended to be restricted to almost mechanical transactional tasks in which one transfers information from one text to another (sometimes across, sometimes within languages/registers).

With descriptors for mediating concepts in collaborative interaction, for goal-oriented (online) collaborative interaction, interaction and mediation strategies, as well as for plurilingual comprehension and the mobilisation of a plurilingual repertoire, the CEFR Companion Volume (CEFR-CV) provides a pointer for a possible future direction for the language assessment of digital natives and the assemblages, code alternation and plurilanguaging that is a natural part of online communication.

Brian North has been a teacher, project coordinator, course director, development manager and researcher and has collaborated extensively with the Council of Europe. He developed the levels and descriptors for the CEFR in his PhD (The Development of a Common Framework Scale of Language Proficiency (2000: Peter Lang)). He co-authored the CEFR, the prototype European Language Portfolio, the Manual for relating examinations to the CEFR, and Eaquals' CEFR Core Inventories (English and French). More recently he coordinated the 2013–2020 Council of Europe project that produced the CEFR Companion Volume and is main editor of the case studies related to it: Enriching 21st century Language Education: The CEFR Companion Volume, Examples from Practice (forthcoming, Council of Europe). Other recent publications include: The

CEFR in Practice (2014: Cambridge Assessment/Cambridge University Press); Language Course Planning (with M. Angelova, E. Jarocsz & R. Rossner: 2018: Oxford University Press), and The Action-oriented Approach (with E. Piccardo, 2019: Multilingual Matters).

Inclusivity issues in language assessment

Ron Zeronis, Cambridge Assessment English

As language testers, we all recognise how important it is to ensure that the material we put in front of learners and teachers reflects the experiences and identities of all the cultures and social groups we serve. Indeed, we all agree that the elimination of bias in our content and the ways in which we deliver our tests is fundamental to fair and ethical testing. And while established conceptual models such as Weir's socio-cognitive framework have helped testers address these issues from a validity point of view, it's only more recently that the impact of equality, diversity, inclusion and belonging (EDIB) in language assessment has become an explicit part of the conversation. Social and racial justice movements like Black Lives Matter have been instrumental in bringing discussion of EDIB issues into the mainstream. And of course, Covid-19 has had a further impact on how our tests and assessments can be accessed, and who can access them.

This workshop will be an opportunity for us as assessment professionals to discuss these issues in some depth, starting from the point of view that:

- our content has impact. The way that we represent people in our assessments has the power to affect the way that our learners see themselves in the world, how they see each other, and how the world sees them.

- our learners include people from every background – ethnic, cultural, linguistic, social, economic, religious – and from across the spectrum of gender identity, sexual orientation, physical ability, neurodiversity, political belief, and personality.

- Some of the key questions to be addressed are:

- What are the practical ways in which we can identify how our content might create or perpetu-

ate systematic bias or advantage, or contribute to prejudice?

- In trying to create content that is authentic, inclusive and representative, how can we make sure that our content does not ignore or marginalise any group, that diverse groups are represented, and that representation is authentic?

- How can we ensure we use our influence as leaders in our field to challenge prejudice and discrimination?

- Do specific language testing contexts (e.g. the language being tested, or the sector or domain of the target users) bring additional or unique challenges?

- What effect has the pandemic had on fairness and accessibility and how can we meet those challenges in a world where how people access and use tests and assessments has been profoundly impacted?

Ron Zeronis is the Head of Content Creation at Cambridge Assessment English. His team works closely with Cambridge Assessment English product development and research teams to provide innovative solutions to test content production at Cambridge Assessment English. He has over 17 years' experience in the Assessment unit at Cambridge Assessment English and holds an MA in Applied Linguistics from the University of Leicester.

Standard setting in the Covid world

Jane Lloyd, Cambridge Assessment English

The practice of standard setting is an established method that language testing organisations use to determine cut scores, pass marks and other grading decisions for their tests. In the last two decades the process has also been used to determine alignment against internationally recognised frameworks such as the CEFR. However, the process is resource heavy and typically involves a group of panellists working together in a room for a long period of time. The Covid world of lockdowns, travel restrictions and an increase in technology use has meant that current ways of carrying out standard setting are no longer possible. Based on a case study of an online standard setting procedure carried out since autumn 2020, this workshop will encourage participants to think how these procedures can still

be implemented using digital means in remote and distance settings.

Jane Lloyd's current responsibilities involve working on large-scale educational reform projects with ministries and NGOs, and advising other national test providers on test validation, statistical analysis, impact research, curriculum reform and quality assurance procedures. Before joining Cambridge Assessment, Jane worked in Japan, South East Asia, Turkey and the Middle East as a teacher, trainer and manager in state schools and higher education for over 20 years. Her interests include alignment to the CEFR and other external frameworks, validation and standardisation of tests and curricula at national level, assessment literacy training for non-specialists, and research projects which involve national organisations, ministries or institutions involved in large-scale testing.

What challenges might there be for language assessment in the 2020's?

Lynda Taylor, Centre for Research in English Language Learning and Assessment, University of Bedfordshire

The first year of the 2020s will undoubtedly be remembered as the Year of Covid-19 – the year when a virus that emerged in one part of the world spread steadily throughout our 'global village' to become a worldwide pandemic, with unforeseen and unimaginable consequences that will have long-lasting effects well beyond the current decade. Like everything else, language testing and assessment across the world suffered unprecedented disruption.

It would be easy to see some of the challenges facing language assessment over the coming decade as resulting directly from the Covid-19 pandemic. In reality, of course, some of these predate the global pandemic and many of us who are involved in language testing and assessment were aware of them well before January 2020.

Three specific areas of longstanding and continuing challenge for language testers relate to: i) our understanding and characterisation of oral communication skills; ii) the role and impact of technology in language test design, delivery and scoring; and iii) the importance of assessment literacy in enabling language tests to contribute positively to the lives of individuals and societies.

This short presentation will consider selected aspects of all three areas of challenge to reflect on how we might conceptualise and address them over the coming decade. We shall consider to what extent the events of the past year have impacted on our perspectives and our resources going forward. How far have the challenges been exacerbated or become more complex due to what happened in 2020? Or is it possible that some challenges are now – or will become – easier to address and to resolve over the coming decade, precisely because of the experience of the past year?

Lynda Taylor is Visiting Professor in Language Assessment at CRELLA (Centre for Research in English Language Learning and Assessment), University of Bedfordshire, UK. Her research interests include issues in speaking assessment, the use of qualitative research methodology, and ethical issues, such as assessment literacy and testing accommodations. She has provided expert assistance for language test design, development and revision projects in the UK and overseas. Key publications include the edited volumes Examining Speaking (2011) and Examining Listening (2013) as well as contributions to peer-reviewed handbooks and journals, e.g. Handbook of Second Language Assessment (2016), Language Assessment Quarterly (2018). She is currently President of the UK Association of Language Testing and Assessment (UKALTA).

GDPR and automarking

Ian Hanahoe, Cambridge Assessment English

Automatic marking of exams not only reduces the burden on schools and awarding bodies but can also greatly speed up the assessment process. However, the GDPR includes strict rules on the use of automatic decision making which need to be carefully considered when employing such tools. This session is designed to give delegates an understanding of what the GDPR says about automatic decision making, when it is allowed and what constitutes human intervention and therefore is not considered a solely automatic decision.

Ian Hanahoe joined Cambridge Assessment in November 2019 as the Data Protection Officer. He has overall responsibility for ensuring Cambridge Assessment is in compliance with data protection legislation around the world through the use of Data Protection Impact Assessments, training of staff, dealing with Subject Access Requests, and providing advice and guidance to colleagues. Before joining Cambridge Assessment, he spent 18 years at the University of Hertfordshire (UH). He supported UH in a number of different legal areas including copyright, intellectual property, corporate law and commercial law he also acted as the Legal Advisor for one of the university's subsidiaries. In 2011 Ian was appointed the Information Compliance Manager for the university dealing with data protection and freedom of information requests. In the lead up to the introduction of the GDPR in 2018, Ian spearheaded the university's preparation for the change in legislation to ensure compliance was maintained, and in March 2018 was appointed the Group Data Protection Officer.

Let's give teachers a hand

Marie-Therese Swabey

Let's give a big hand to teachers for their flexibility and resourcefulness! What a year it's been and haven't teachers been fantastic?! All over the world teachers turned on a sixpence and moved their classes online, many with little or no training for online teaching. It's been a steep learning curve for us all and we've learnt a lot about technology and about teaching online.

In this workshop, we'll start with a review of some interesting technology teachers can use online to support learning and prepare learners for exams. For example, using the features of different platforms, such as Zoom and Teams, and using digital tools, such as Padlet and Write and Improve, to help learners with listening, speaking and writing tests or using Quizlet for reading, vocabulary and grammar work.

Then we'll talk about next steps and ways to move forward now that we know that online learning is here to stay. There will be a challenge for language testers attending the workshop to think about and discuss ways that they can look after teachers who prepare students to take their exams. We'll focus on things teachers say they are concerned about when preparing learners for exams in online lessons. For example, accessing materials and resources to prepare learners for tests; planning different types of lessons for different tests; keeping online learners motivated; dealing with connection issues; setting tests; trying to replicate exam conditions for learners who are at home; managing practice tests' marking and grading learners' work and assessing learners.

We'll share ideas, experiences and possible solutions to teachers' problems so that they are better able to support learners to be successful in their exams.

Marie Therese Swabey has been involved in English language teaching for more than 30 years as a teacher and a teacher trainer. She is a Joint

Chief Assessor for Cambridge CELTA courses providing support to new and existing centres. She is Chief Moderator for CELT-S/CELT-P courses and an item writer for the Cambridge Assessment English TKT and OET exams. She has written materials for teachers and teacher trainers for Cambridge Assessment English and Cambridge University Press, including the Cambridge University Press Grammar for Teachers online course and the Cambridge Assessment English Language for

Teaching and CELT-S/CELT-P courses. More recently she has been involved in the development of Cambridge Assessment English online courses for teachers on Future Learn including: Exploring the World of English Language Teaching, Teaching English Online, and the Cambridge English Train the Trainer course. She has been the Lead Educator on a number of these courses working with teachers and teacher trainers from around the world.

Uneven profiles: What it is and why it is important

Lorenzo Rocca, ALTE and Council of Europe expert member

Cecile Hamnes Carlsen, Western Norway University of Applied Sciences

Brian North, Eurocentres Foundation

This workshop focuses on the topic of uneven profiles, a key concept in the CEFR Companion Volume (2020). It is quite natural for language users, first and second language users alike, to have somewhat different levels of proficiency in the different language skills. Many of us can read Shakespeare, few of us can write like him. For those migrants who for different reasons have limited literacy skills, it is common to perform better in the oral skills (listening and speaking) than in the written skills (reading and writing). To give all candidates a fair chance in testing, measuring the different skills in different parts of the test yielding independent scores is therefore an advantage. Many test developers do exactly that. A related question of equal importance, then, is to what extent those who set formal language requirements, like migrations policy makers and employers, set differentiated requirements, demanding higher levels in some skills than in others. There is reason to assume that this possibility is not fully exploited in practice.

This workshop opens with a general introduction to the notion of uneven profiles, given by Brian North, one of the authors of both the CEFR and the CEFR Companion Volume. His introduction will respond to the questions of what uneven profiles are and why it is important to take a differentiated profile approach in assessment.

Following the introduction, workshop participants will be divided into two groups according to their interests. Group 1, chaired by Lorenzo Rocca, will focus on the topic uneven profiles and low literate migrants. Participants will work on the basis of two overall scales from the LASLLIAM reference guide for developing two adaptive tasks for low-literate language learners. Group 2, chaired by Cecilie Hamnes Carlsen, will focus on uneven profiles and language requirements within the migration context. Participants discuss issues related to the practical development of listening tasks within a differentiated profile approach.

The workshop will round off with a round table summing up insights from the two working groups.

Lorenzo Rocca has a degree in Classics and a postgraduate degree in the Didactics of L2 Italian. After having been a teacher of Italian as a Foreign Language, he worked at the CVCL (Centre for Evaluation and Language Certification) within the University for Foreigners of Perugia. His duties ranged from marking and examining to item writing and running seminars for teacher training. Since 2006 he has been responsible for research projects at national and supranational level, focused on connections between teaching and assessment in the migration context; within this focus he has published articles, syllabi for language provision and specifications for CVCL exams for adult migrants. He has been the chair of Language Assessment for Migrants' Integration (LAMI) in Association of Language Testers in Europe (ALTE) since 2008 and a member of the Council of Europe LIAM project since 2014. In 2019, he assumed the overall responsibility for the activities of the CVCL and in 2020 he was awarded with the national scientific qualification as Associate Professor. He is currently finalizing a long-term project with the Council of Europe – a reference guide for low-literacy migrant learners.

Cecilie Hamnes Carlsen is professor in second language acquisition at Western Norway University of Applied Sciences where she teaches and supervises on the topics second language acquisition, language assessment, language policy and multilingualism. Since 1997 she has worked mainly in language testing and validation projects for adult migrants, a work she was in charge of from 2011 to 2017. Her research interests cover topics related to the use and misuse of language tests in the migration context, in the labour market and in education, with a particular focus on the impact of language tests on low-literate adult learners. She is an expert member of ALTE, member of the Standing Committee and co-chair of the LAMI project group.

Brian North has been a teacher, project coordinator, course director, development manager and researcher and has collaborated extensively with the Council of Europe. He developed the levels and descriptors for the CEFR in his PhD (The Development of a Common Framework Scale of Language Proficiency (2000: Peter Lang)). He co-authored the CEFR, the prototype European Language Portfolio, the Manual for relating examinations to the CEFR, and Eaquals' CEFR Core Inventories (English and French). More recently he coordinated the 2013–2020 Council of Europe project that produced the CEFR Companion Volume and is main editor of the case studies related to it: Enriching 21st century Language Education: The CEFR Companion Volume, Examples from Practice (forthcoming, Council of Europe). Other recent publications include: The CEFR in Practice (2014: Cambridge Assessment/Cambridge University Press); Language Course Planning (with M. Angelova, E. Jarocsz & R. Rossner: 2018: Oxford University Press), and The Action-oriented Approach (with E. Piccardo, 2019: Multilingual Matters).

Measuring cognitive processes in assessment contexts – the case of eye-tracking

Elaine Schmidt, Cambridge Assessment English

Research in language assessment and testing has largely concentrated on analysing output, i.e. learners' writing or speech, or results of multiple-choice or written answers in listening and reading tasks. While discussions of cognitive validity of assessment tasks have become more frequent over the last few years, actual experimental data looking at cognitive processing are sparse. Eye-tracking can be used to tap into underlying cognitive processes which questionnaires and other written or oral responses cannot capture since these are by necessity influenced by deliberation. Furthermore, the few studies which have used eye-tracking tools in assessment contexts often analysed the data in a qualitative way by looking at retrospective think-aloud reports from participants. Thus, despite using a tool to investigate underlying cognitive processing that most readers are unaware of, many studies have focused on using eye-tracking to analyse subjective deliberations and attention-driven bottom-up processes (which also require the studies' subjects to display a certain level of meta-awareness).

This talk provides an overview of eye-tracking methodologies, explaining how they work, their limitations and the value they have added to the language processing and learning research of first and second language learners. Finally, we will discuss the possibilities and potential of using eye-tracking in a language assessment context, specifically in the evaluation of processing load and the development of test materials.

Elaine Schmidt's research focuses on cognitive aspects of language processing and learning using eye-tracking and electroencephalography (EEG). She obtained her PhD in phonetics and bilingual language acquisition from the University of Cambridge. After her PhD she worked on cognitive processes of L1 and L2 speech perception in Sydney, Australia, and again back at Cambridge. Half a year ago, Elaine then joined Cambridge Assessment where she brings her expertise in eye-tracking and EEG in second language learning to an assessment context.

Optimiser l'arbitrage grâce à la notation automatique

Dominique Casanova, Le français des affaires, CCI France Education

Depuis plusieurs années, on constate une tendance à la numérisation des tests de langue, et la pandémie mondiale a renforcé cette tendance. Il devient en effet compliqué, voire impossible, de rassembler de grands nombres de candidats dans une même pièce ou d'échanger physiquement des livrets de tests avec les candidats, voire simplement de les acheminer. Les tests au format numérique offrent la possibilité aux centres de passation d'organiser des sessions multiples, à effectifs limités, en se jouant des frontières. La surveillance à distance permet même, dans certains contextes, d'organiser des tests au domicile des candidats.

Au-delà de ces aspects organisationnels, les tests au format numérique permettent de récupérer des informations supplémentaires sur la passation des épreuves, comme le temps passé sur chacune des questions ou la dynamique de saisie des productions écrites. La passation au format numérique des épreuves d'expressions permet aussi de constituer aisément des corpus de productions de candidats, qui peuvent dès lors être exploités à des fins de recherche ou d'amélioration de la qualité des tests.

Ainsi, le développement des outils et recherches en traitement automatique des langues et en intelligence artificielle rendent relativement aisé le développement de systèmes de notation automatique, au-delà de la langue anglaise. Certes, de tels outils doivent encore faire leur preuve et la question de la validité de leurs résultats reste posée. Mais on peut dès à présent leur trouver une place aux cotés des humains. Ils peuvent, par exemple, servir de garde-fou pour déclencher des contrôles d'évaluations ou permettre d'automatiser certaines décisions d'arbitrage en cas d'écart entre évaluateurs.

Dans cette présentation nous exposerons les travaux menés par Le français des affaires pour développer un système de notation automatique des épreuves écrites de tests de langue française et l'utilisation qui est en-

visagée de cet outil pour renforcer le monitorage des évaluations et la fidélité des scores délivrés au test.

Dominique Casanova est responsable du développement scientifique au Français des affaires, établissement du groupe CCI France Éduca-tion. Ingénieur en informatique et mathématiques appliquées, il a vécu dix ans au Maroc où il a notamment dirigé les cours de langue de l'Insti-tut français de Rabat. Coprésident du groupe de travail Technology in Language Assessment au sein de l'association ALTE, ses principaux centres d'intérêt sont la psychométrie et l'utilisation des technologies numériques dans l'évaluation, notamment le traitement automatique des langues et l'intelligence artificielle. Ses contributions scientifiques sont disponibles à l'adresse https://www.researchgate.net/profile/Domi-nique_Casanova/contributions

Helping or hindering? The opportunities and challenges of assessing interactional skills using video-call technology

Evelina Galaczi, Cambridge Assessment English

Hye-won Lee, Cambridge Assessment English

Mina Patel, British Council

How can we harness technology to enhance the development and assessment of interactional skills?

There is no easy answer to this question since technology can have both an expansive and a reductive impact on speaking assessment. It can support new task types and modes of delivery, but it can also narrow the test construct (Chapelle and Douglas 2006). To leverage technology to the benefit of the learner, we need to look beyond speaking tests that only rely on one-way mono-logic responses to computer-delivered prompts. We need to think through the possibilities and the pitfalls of using technology to enhance tests of interactional competence.

This presentation will explore how we can use technology in speaking assessment without detriment to the interactional nature of tests. We will start with an overview of the evolution of the interactional competence construct and the current state of knowledge about and understanding of this construct, followed by the discussion of some of the key challenges in assessing interactional competence. We will then turn to the role of technology in the assessment of interactional competence and present a research study of a video-call speaking test used in an international high-stakes environment. A four-phase programme of work involving 595 test-takers in seven international locations investigated video-call speaking tests and how similar or different they are from in-person face-to-face tests. We will present data on test scores, language functions and ex-aminer and test-taker perceptions, and analyse the key similarities and differences between in-person and video-call test modes.

The session will conclude by looking at the funda-mental questions that need to be addressed empirically as part of the validation of video-call speaking tests, and the insights such empirical investigations can offer for the emerging construct of online interactional com-petence.

Evelina Galaczi is Head of Research Strategy at Cambridge Assess-ment English. She has worked in language education for over 25 years as a teacher, teacher trainer, materials writer, researcher and assess-ment specialist. Her areas of interest and expertise are in speaking as-sessment, the integration of digital technologies with assessment and learning, and developing assessment literacy. She holds Master's and Doctorate degrees in Applied Linguistics from Columbia University, USA.

Hye-won Lee is Senior Research Manager at Cambridge Assessment English where she conducts research projects related to new generation assessments and new products. Before joining Cambridge English, she gained extensive experience developing and validating digital assess-ments at leading organisations based in the USA. She holds a PhD in Ap-plied Linguistics and Technology from Iowa State University. Her current work focuses on the use of video call technology in speaking tests and the proficiency model of language ability in data-driven diagnostic as-sessment.

Mina Patel is Assessment Research Manager with the Assessment Re-search Group at the British Council. Her background is in English lan-guage teaching and training. She has worked in the UK, Greece, Thail-and, Sri Lanka and Malaysia as a teacher, trainer, materials developer and ELT projects manager and has extensive experience working with Ministries of education in East Asia. Her interests in language testing and assessment lie in the areas of language assessment literacy and the impact of testing and assessment. She is currently a PhD student with CRELLA at the University of Bedfordshire, UK.

1ST INTERNATIONAL DIGITAL SYMPOSIUM
#ALTEdigital2021

'I teach in a non-traditional way so I need non-traditional assessment': Teachers' assessment beliefs and practices in virtual exchange

Anna Czura, Autonomous University of Barcelona

Melinda Dooly, Autonomous University of Barcelona

Telecollaborative or virtual exchanges (VE), entail the 'engagement of groups of students in online intercultural exchange, interaction and collaboration with peers from partner classes in geographically distant locations, under the guidance of educators and/or expert facilitators' (INTENT project; Position paper; Virtual Exchange in the European Higher Education Area, 2014). VE often takes place in a multilingual setting, in which students use two or more languages to communicate. In light of students' limited mobility, particularly visible in the time of the pandemic, online intercultural exchange at schools and tertiary levels contributes to broadening the scope of internationalisation at home and internationalisation of the curriculum. Despite the growing popularity of VE projects in FL education around the world, there is an acute shortage of empirical evidence and practical publications alike on the topic of assessment in this mode of education. As an inherent element of any VE project, assessment poses a considerable challenge as it is 1. culture-dependent and, hence, largely determined by an educational and institutional context; 2. multifaceted – embracing not only linguistic but also intercultural and technical components; and 3. of dynamic, interactive nature, and hence unpredictable.

In this presentation, we provide an overview the ASSESSnet project – a mixed-methods research project which aims to explore teachers' beliefs about assessment and analyse the objectives, process and content of foreign language assessment in virtual contexts at tertiary level. Our main emphasis will be placed on presenting the results of the questionnaire and interviews with teachers that focused on their current assessment practices and assessment tools. The findings will be used to open up discussion regarding promising practice and potential training resources and practical recommendations regarding VE assessment.

Anna Czura is a MSCA-IF Postdoctoral Researcher at the Education Faculty of the Universitat Autònoma de Barcelona (Spain). Her teaching and research interests centre on foreign language assessment, telecollaboration at tertiary level, teacher education, student teaching abroad and CLIL. She is an expert for the Ministry of Education in Poland and a member of the GREIP (Research Center for Plurilingual Teaching & Interaction) research group.

Melinda Dooly holds a Serra Húnter fellowship as teacher and researcher at the Education Faculty of the Universitat Autònoma de Barcelona (Spain). She teaches English as a Foreign Language Methodology (TEFL) and research methods courses, focusing on telecollaboration in education at both undergraduate and graduate levels. Her principal research addresses technology-enhanced project-based language learning in teacher preparation. She is co-editor of the book series Telecollaboration in Education (Peter Lang), with Robert O'Dowd.

Portuguese as a Foreign Language in China: testing challenges in future regarding contexts, profiles and pandemic restrictions

Catarina Gaspar, University of Lisbon

Maria José Grosso, University of Lisbon

Milena Sousa, CAPLE

The teaching and learning of Portuguese as a Foreign Language (PFL) in China is growing with an increasing number of universities and institutions offering Portuguese language courses integrated (or not) in degree tracks, master and/or doctorate ones, the latter provided mainly by Macao and by some institutions in China. Although the Portuguese language is a 'small language', when compared to other languages such as English, it is also a strategic and attractive one for many students looking for professional opportunities in the globalised routes that cross the Portuguese speaking countries, like Brazil, Angola or Mozambique, among others. First, we intend to discuss if this context demands the growth of the available places to apply for the PFL international certificates. Moreover, our aim is to identify the challenges of testing PFL in China, concerning the academic, technological and sociocultural contexts, and the profile of Chinese students. In 2020-2021, the COVID-19 pandemic situation posed new

challenges to the PFL testing for the Chinese audience due to travel restrictions and the suspension of international academic mobility programs. We aim to make a preliminary analysis of how this affected the PFL testing. Thus, this paper intends to provide a first overview of the challenges and new directions that testing of the PFL should take in the future in China, namely, how can new technologies offer solutions to fulfil the assessment and learning needs of this audience and guarantee that PFL testing is in line with PFL teaching and learning.

Catarina Gaspar teaches in the University of Lisbon (UL) since 2000; she works in the Post-Graduation in PFL/PL2 and Director of the MA in PFL/L2. In the past, was sub-director in CAPLE (2009-2013). She holds a PhD in Latin Linguistics (2009). Among her research interests are sociolinguistics, language assessment, and learning and teaching. She presented papers in international conferences and is the author of articles and book chapters. Now she is in charge of the seminar 'Teaching, Learning and Assessment'.

Maria José Grosso is Associate Professor in the University of Lisbon, and since 2012, she is Invited Professor in the University of Macau. She was Director of the CAPLE (2009-2012). She holds a PhD in Applied Linguistics (2000). Her areas of study and teaching focus on Applied Linguistics to the Teaching of Portuguese (FL/L2) and benchmarks in the areas connected to it, including teaching, learning, assessment (FL/L2), disclosure policy and language teaching and intercultural studies.

Milena Sousa is researcher and senior office of CAPLE since 2002. She has bachelor in Linguistics (2002) and Post-graduation in teaching PL2. She has been working in language testing for 15 years in Test Administration, test construction, item writer, test analysis; communication with exam centres and general stakeholders. She is also teacher of PFL in Universidade de Lisboa, teaching Erasmus language courses and PFL general courses.

Automarking as a supplementary tool for arbitration between raters

Dominique Casanova, Le français des affaires, CCI France Education

In recent years, there has been a trend towards the digitalization of language tests, which has been reinforced by the global pandemic. It has become complicated, if not impossible, to gather large numbers of candidates in the same room or to physically exchange test booklets with candidates, or even simply to ship test material. Digital tests offer the possibility of avoiding having test material cross international borders as well as the need for physical contact. Remote proctoring even makes it possible to organize at-home tests for candidates.

Beyond these organizational aspects, digital tests make it possible to collect additional information during test-taking, such as the time spent on each question by a candidate or the dynamics of his typing during writing tests. Offering digital tests for writing or speaking also makes it easy to build up corpora of candidate productions. Such corpora can then be used for research purposes or to improve the quality of the tests.

The development of tools and research in natural language processing and artificial intelligence make it feasible to develop automatic rating systems beyond the English language. Of course, such tools have yet to prove themselves and the question of the validity of their results remains open. But they may find a place alongside humans. They can, for example, serve as a safeguard to trigger rating controls or allow the automation of some arbitration decisions in the event of a discrepancy between raters.

In this communication we will present the work carried out by Le français des affaires to develop an automatic scoring system for written examinations in French language tests and the intended use of this tool to strengthen the monitoring of ratings and the reliability of test scores.

Dominique Casanova is the head of scientific development at Le français des affaires/French for Business, an establishment of the CCI France Education group. An engineer in computer science and applied mathematics, he lived in Morocco for 10 years, where he notably managed the French Institute's language centre in Ribat. He is Co-chair of the Technology in Language Assessment SIG group within ALTE, and his main areas of interest are psychometrics and the use of digital technologies in assessment, notably natural language processing and artificial intelligence. His scientific contributions are available at https://www.researchgate.net/profile/Dominique_Casanova/contributions

Social justice and structural racism: the role of language, language tests and language testers

Cecilie Hamnes Carlsen, Western Norway University of Applied Sciences

Over the past decades, the professional field of language testing and assessment has made significant advancements. Methodological improvements, like Item Response Theory (IRT) and multi-faceted Rasch analysis, enable language testers today to largely overcome challenges of rater inconsistency, item bias and lack of consistency across test administrations – all severe threats to test-related fairness (McNamara and Knoch 2012). Parallel to these professional improvements, however, we have witnessed an increasing use of language test scores as the basis of very high-stakes decision regarding an individual's human and democratic rights – such as the right to residency, family reunification, citizenship, labour, education and housing (Bruzos et al 2018, Rocca et al 2020). Hence, while professional language testing has overcome problems related to fairness, challenges related to justice are more pressing than ever.

In this presentation I explore the concept of structural racism (Bangstad 2021, Midtbøen 2021) as a barrier to social justice in the migration context. After an introduction to this concept and its relevance, I turn to the main question addressed in the presentation: what role does language, language tests and language requirements play in structural racism? And moreover, what is the responsibility of professional language testers in preventing test scores influencing policies and practices in a way that leads to social injustice? In the presentation, I argue that language testers in a Post-Covid world have a professional responsibility to address test misuse and social injustice, the persisting challenge facing our field today. How this can be done in practice is illustrated by reference to the language test activism model (Carlsen and Rocca, forthcoming).

Cecilie Hamnes Carlsen is professor in second language acquisition at Western Norway University of Applied Sciences where she teaches and supervises on the topics second language acquisition, language assessment, language policy and multilingualism. Since 1997 she has worked mainly in language testing and validation projects for adult migrants, a work she was in charge of from 2011 to 2017. Her research interests cover topics related to the use and misuse of language tests in the migration context, in the labour market and in education, with a particular focus on the impact of language tests on low-literate adult learners. She is an expert member of ALTE, member of the Standing Committee and co-chair of the LAMI project group.

Hybrid marking – combining human examiners and automated assessment

Edmund Jones, Cambridge Assessment English

Mark Brenchley, Cambridge Assessment English

The last decade has seen a marked increase in the development and use of automated assessment systems in educational settings. One of the central considerations regarding such systems is the extent to which they can be used without compromising the reliability of the scores given to candidates.

In this talk, we will provide an overview of one response to this consideration. This is the 'hybrid marking' model that Cambridge English is now using to mark the Linguaskill Speaking exam, a computer-delivered oral English test. In operation since July 2020, this model determines whether the score that a candidate receives should be that of a certified examiner or that of the Linguaskill Speaking automarker. If the model is confident that an examiner would give an equivalent (i.e. sufficiently similar) score to the automarker, then the candidate is awarded the automarker score; otherwise, the response is instead passed to an examiner and the candidate is awarded the examiner's score.

We will discuss the principles and operational mechanics of the model, focusing on how the model is designed to set a minimum standard that ensures robust assessment while also allowing for further improvements in the underlying automarker. We will describe our core evaluation criteria, explaining how they differ from other automarking approaches, and provide evidence for the effectiveness of the model. We will also offer some details of the Linguaskill Speaking automarker itself, known as 'CASE', discussing how it is designed and developed as an NLP system.

Edmund Jones holds a PhD in computational statistics from Bristol Uni-

versity. As a Senior Research Manager at Cambridge Assessment English, he manages research related to computer-based tests, automated assessment, and detection of malpractice. He provides expertise in quantitative research methods for benchmarking and impact assessment projects for governmental education ministries. Previously he worked in the Department of Public Health and Primary Care at Cambridge University on statistical modelling of cardiovascular disease. He has also worked in Japan as an English teacher and a translator.

Mark Brenchley holds a PhD in education from the University of Exeter. As a Senior Research Manager at Cambridge Assessment English, he manages research in the areas of writing, lexico-grammar, and auto-marking systems. He has a special focus on the development and application of corpora and corpus-based methodologies, managing both the Cambridge Learner Corpus and the Cambridge English Profile Corpus. For his PhD, he explored the development of spoken and written syntax in L1 English students. Following his PhD, he co-developed the Growth in Grammar Corpus, a novel corpus of student writing that covers the primary and secondary phases of the English education system. He has also taught in the English education system across a variety of contexts for many years, both mainstream and non-mainstream.

Impacto del COVID en el presente y el futuro de los exámenes de dominio del Instituto Cervantes

Javier Fruns Gimenez, Instituto Cervantes

Ines Soria Pastor, Instituto Cervantes

Como a todas las instituciones certificativas, el COVID-19 ha obligado al Instituto Cervantes a repensar la forma de administración de sus exámenes de dominio. Las decisiones se han tomado en función de las características de cada examen, tanto con respecto a cómo se administraban como a la naturaleza del certificado que proporcionan.

- Diplomas de español DELE. Los exámenes para la obtención de los DELE se organizan en convocatorias masivas en centros de examen autorizados y se administran en papel. Durante 2020 se han implementado las siguientes medidas: Administración de las pruebas orales por ordenador con el candidato en el centro de examen; Proyección del material de las pruebas de expresión e interacción orales.

- Servicio Internacional de Evaluación de la Lengua Española (SIELE). SIELE es un examen multinivel por ordenador, que se administra mediante un sistema de citas individuales en centros de exa-

men autorizados. A raíz de los confinamientos, se ha desarrollado un sistema de administración en remoto para evitar el desplazamiento a los centros de examen.

- Prueba de conocimientos constitucionales y socioculturales de España (prueba CCSE). Aunque no se trata de un examen de dominio lingüístico, constituye un requisito para solicitar la nacionalidad por residencia y es elaborado por el Instituto Cervantes. Se trata de una prueba 25 ítems de opción múltiple seleccionados de un banco de 300 ítems. Debido a la crisis sociosanitaria, se ha configurado un sistema de administración por citas y de administración por ordenador y en remoto.

En la ponencia analizaremos estas medidas y otros posibles desarrollos en el futuro.

Javier Fruns Giménez es Responsable de la Unidad de Certificación Lingüística del Instituto Cervantes. Ha participado en el diseño y la elaboración de los exámenes DELE y SIELE, así como en la elaboración de los cursos para examinadores. Regularmente imparte cursos sobre evaluación de lenguas en universidades y escuelas oficiales de idiomas. Javier es doctor en Letras por la Universidad de Massachusetts y licenciado en Filología HIspánica.

Inés Soria Pastor es Jefa del Departamento de Certificación y Acreditación del Instituto Cervantes. Ha coordinado el diseño de los exámenes DELE vinculados al MCER y ha dirigido la creación de los cursos para examinadores. Inés ha liderado el equipo del Instituto Cervantes en el desarrollo del examen SIELE y ha participado en la puesta en marcha del sello SICELE. Es miembro del Grupo de Interés en la Evaluación de Lenguas en España. Inés es máster en Enseñanza de ELE por la Universidad Nebrija y licenciada en Filología Hispánica.

Adaptive testing: New challenges in assessment. Case study of esPro

Marta García, Universidad de Salamanca

Yiannis Papargyris, LanguageCert

LanguageCert USAL esPro is a manifestation of the inextricable link between assessment and technology. Although the test was originally developed as a paper-based test, it is now available as a computer-adaptive, online-proctored exam. In this talk, we trace the challenges we had to overcome in order to ensure that the recently developed adaptation maintained the long-established principles of assessment development.

Marta García is a Spanish as Foreign Language teacher in Cursos Internacionales de la Universidad de Salamanca and responsible for the LanguageCert USAL esPro test. She holds a BA in German Studies, a Master's Degree in Teaching Spanish as a Foreign Language and a Doctorate in Spanish Philology, all of them from the University of Salamanca.

Yiannis Papargyris works for LanguageCert as Language Assessment Development Manager He holds a PhD from the University of Birmingham and is a Chartered Educational Assessor.

Language educators working online during the COVID-19 pandemic: Learning, teaching and assessment practices in the limelight

Dina Tsagari, Oslo Metropolitan University

Jenny Liontou, National and Kapodistrian University of Athens

Christina Nicole Giannikas, CG Education & Research Consultancy

The abrupt global outbreak of the Covid-19 pandemic in early 2020 led to emergency remote teaching (ERT) in all fields of education. In the blink of an eye, educational authorities worldwide called for school closures to limit the spread of the pandemic, a situation that prompted UNESCO to urge countries to ensure flexible learning in formal and informal settings (Huang et al 2020). The abrupt shift to online teaching came as a shock to the majority of the language education community who had to become familiar with online tools, resources, and adjust their teaching approaches overnight (Giannikas 2020). At the same time, while teaching is moving online on an untested and unprecedented scale, teachers did not necessarily have the technologies, resources, knowledge and strategies needed to succeed in emergency online courses (Xie et al 2019). This situation presented a need for empirical research to address the unique features of learning environments (e.g., physical context, social cultural context, technology context, instructional context, etc.) as well as the unique characteristics of teachers and learners (Bolliger and Martin 2018, Martin et al 2018) in these contexts during this quick transition to emergency online learning. Given the circumstances, the present paper presents the findings of an international survey that aimed at providing valuable insights into online foreign language teaching practices in the European education system by investigating

whether and to what extent the aforementioned context actually facilitated and fostered students' engagement while focusing on how to design learning environments, pedagogies, and strategies to engage and assess learners during and after this quick transition to online learning. Our findings are expected to shed light on the struggles the ELT community faced in order to reimagine asynchronous and synchronous teaching in a digital learning environment, and refer to the impact and lessons learned from the global experience of ERT.

Dina Tsagari, PhD, is Professor at the Department of Primary and Secondary Teacher Education, Oslo Metropolitan University, Norway. She has also worked for the University of Cyprus, Greek Open University and Polytechnic University of Hong Kong. Her research interests include language testing and assessment, materials design and evaluation, differentiated instruction, multilingualism, distance education and learning difficulties. She is the editor and author of numerous books, journal papers, book chapters, project reports etc. She coordinates research groups, e.g. CBLA SIG – EALTA, EnA OsloMet and is involved in EU-funded and other research projects (e.g. DINGLE, ENRICH, TALE, DysTEFL, PALM, etc.)

Jenny Liontou, PhD, is Assistant Professor at the Department of Language and Linguistics, Faculty of English Language and Literature of the National and Kapodistrian University of Athens, Greece. Specialising in the area of Applied Linguistics, ELT methodology and assessment practices, she has made presentations in national and international conferences and has published papers in the aforementioned fields. Her current research interests include theoretical and practical issues of EFL pedagogy, testing and assessing foreign language competence and corpus-based teaching and assessment practices. Her recent publications include chapters in edited volumes by Cambridge University Press, Routledge, Peter Lang and Cambridge Scholars Press, along with articles in international peer-reviewed journals.

Christina Nicole Giannikas holds a PhD in the field of Applied Linguistics. She is a consultant and founder of CG Education & Research Consultancy. She also works in Higher Education where she lectures courses in Applied Linguistics and is a teacher trainer for the pre-service and in-service teacher education programs in Cyprus and beyond. She specializes in the areas of early language learning, age-appropriate digital pedagogies, digital literacies, assessment, and teacher education. For the last four years, she has been serving as Chair of the EuroCALL Teacher Education SIG and was recently appointed the EuroCALL National Contact for Greece. Keep an eye out for her upcoming new book with the title: *Teaching Practices and Equitable Learning in Children's Language Education*!

Using two dimensions to assess a three-dimensional language: Chances and challenges of online testing for sign language proficiency

Tobias Haug, Interkantonale Hochschule für Heilpädagogik

During the pandemic, we all were forced to move teaching and assessment of sign languages online. This has been done by using video-conferencing tools, learning platforms, or even survey instruments.

In my presentation, I will provide examples how we have moved the assessment of Swiss German Sign Language proficiency from on-site to online. While this change from on-site to online can be done fairly easy for testing receptive skills, it is much harder for productive skills since a number of issues need to be considered to produce video-recordings of good enough quality, for example, good lightning conditions, or a monochromatic and light background. While the learning curve for solving technical issues was very steep in the last 12 months, proctoring for contexts like final exams is still unsolved.

In a post-pandemic world, online assessment of sign language proficiency will remain an important issue but will not replace the assessment of real-life interaction of a sign language. Online assessment in combination with sign language recognition (which can also include an automatic feedback loop) will create new learning opportunities for students.

There are plans underway to document the experiences made in the last year in online sign language teaching and assessment. Documenting the chances and challenges that were made during the pandemic might serve as a basis for future online assessment scenarios.

Tobias Haug studied Sign Linguistics at Hamburg University, Germany, and Deaf Education at Boston University, USA, where he received his master's (Deaf Education) in 1998. In 2009 he earned his PhD (Sign Linguistics) at Hamburg University, Germany. In 2017, he completed a distance Master's in Language Testing from Lancaster University. From 1998 to 2004, he worked as a sign language interpreter and researcher. Since 2004, he has been the director of the sign language interpreter

program and professor in Zurich, Switzerland at the Interkantonale Hochschule für Heilpädagogik (HfH). He has been involved in different international and Swiss research projects investigating the development and validation of tests for Swiss German Sign Language, rater behavior in the rating process, and the use of new technologies in sign language assessment.

Redressing the balance post-Covid – the challenges for the workplace language proficiency assessment: a case study in the aviation environment

Neil Bullock, Lancaster University

Few can doubt that the last 12 months have been challenging for all practitioners in education. Both teaching and testing have had radical changes forced upon them. Teachers have had to develop new methodologies not just for learning but how best to integrate additional skills like technology. Testers have had to rethink their operations dramatically, be it the simple logistical challenges of the testing room to protect candidates and examiners alike, or the more problematic issues of moving to online testing with the obvious fears over test security, rater performance, and the reliability of technology to run the exams.

Language learning and assessment for pilots and air traffic controllers (ATCOs) have not been spared the challenges of the Covid world. An industry massively hit financially and psychologically because of the huge reduction in people travelling has had to ensure that, because of regulatory requirements, staff still remain operationally current, and this includes mandatory language proficiency testing for most of the world's pilots and ATCOs, whatever language they use to communicate.

How much these changes will affect a post-Covid world is still a long way from being known. It has been suggested that learning and testing as we know it will never return to the traditional methods. Other have gone further and advocated that this is the time for a radical rethink in testing language proficiency in aviation that should address many of the issues that have arisen since obligatory testing was introduced in 2008.

This presentation will analyse results of a survey carried

1ST
INTERNATIONAL
DIGITAL
SYMPOSIUM
#ALTEdigital2021

Invited Presentations

out amongst language trainers and testers in the aviation world to see how the enforced changes of the Covid period have provided positive incentives to adapt and improve a system of learning and testing in aviation. It will go on to address whether such changes can, in fact, resolve longer term issues in language testing in this domain, for a positive post-Covid outcome.

Based in Switzerland, **Neil Bullock** is a teacher, teacher trainer and test developer working with Aviation and Business English. He has overseen test and assessment-related projects around the world on behalf of various clients, including: rater training, item writing, assessment literacy workshops for teachers, test centre management, oral examining, and national education policy evaluation. He is also a former Coordinator and webinar moderator for IATEFL Testing Evaluation and Assessment SIG and has been a member of ALTE for almost 10 years. He has an MA in Applied Linguistics, including a dissertation on test development for language proficiency in aviation and is currently in the second year of doctoral studies at Lancaster University researching the assessment of communicative competence in aeronautical communication.

Integrating mediation skills into a computer-based speaking test

Jing Xu and Annabelle Pinnington, Cambridge Assessment English

The Covid-19 pandemic has forced an abrupt switch to home-based working and learning and has significantly changed the way people communicate. Compared to face-to-face communication, online communication is usually more challenging in that it may result in reduced effectiveness of nonverbal communication, difficulty in turn-taking management and communication breakdown due to technical issues such as poor sound quality and delayed video transmission (Nakatsuhara, Inoue, Berry and Galaczi 2017). In virtual meetings, mediation, or an interactive process that is used to facilitate communication, understanding, and/or learning (Council of Europe 2001), has become indispensable. The latest Companion Volume to the Common European Framework of Reference for Languages (Council of Europe, 2018) provides a conceptual scheme for mediation and a blueprint for enriching this construct. Despite this increased recognition of the importance of mediation to oral communication, the construct has rarely been covered in L2 speaking assessment.

This paper reports a study that trialled a number of

computer-based speaking tasks designed to assess mediation. The focus is specifically on a task called 'evidence-based recommendation', which asks a candidate to select and link information from multi-modal sources and make a recommendation based on the evidence presented. The participants of the study were 340 L2 English learners whose oral proficiency ranged from B1 to C2. They were closely observed while performing the tasks and were interviewed on completion of them. Discourse analysis was conducted on the speech elicited by the task. Analysis of variance (ANOVA) was conducted to investigate the differences in task performances across four proficiency groups. The study found that the task elicited four types of mediation behaviours and was useful for distinguishing among B1, B2 and C1 or above speakers. In addition, the participants considered the task authentic, engaging but cognitively demanding.

Jing Xu is a Principal Research Manager at Cambridge Assessment English, which he joined in November 2015. Jing began his career as an English language teacher and before moving to Cambridge he worked as the Program Assistant for the Oral English Certification Test at Iowa State University and a summer research intern at Educational Testing Service. He obtained his doctorate in Applied Linguistics and Technology from Iowa State University. His research focuses on L2 speaking assessment, automated scoring and feedback and computer-assisted language learning. He is currently an Editorial Advisory Board member for the Language Assessment Quarterly journal.

Annabelle Pinnington is a Senior Assessment Manager in Cambridge Assessment English and has been working on test production and development since 2014. Prior to joining Cambridge Assessment, she worked in the UK, Russia, Indonesia and Malaysia as an English language teacher and Academic Manager, specialising in teaching advanced levels and teacher training. Her work in Cambridge Assessment English focuses primarily on the speaking components of computer-based multi-level tests. She is currently the co-chair of ALTE's CEFR Special Interest Group and working on her MA in Educational Assessment at UCL.

Fair and just arrangements for dyslexic test takers?

Hanne Lauvik and Stine Sørlie, Skills Norway

In 2020, Skills Norway was asked by the Ministry of Education and Research to conduct trials of test arrangements for dyslexic candidates in the test in Norwegian for adult immigrants and consider revisions in the arrangements offered to dyslexic candidates in the

test. This initiative came from adult education centres and organisations working to improve conditions for those struggling with dyslexia, who were concerned that these test takers would not be able show their full language potential without access to assistive tools in the reading and writing part of the test.

Skills Norway set up the trials in which adult candidates with dyslexia had access to assistive technology tools. The objective of the trial was to gain knowledge about how these aids might offer a better assistance to dyslexic test takers and to put us in a better position to evaluate whether digital aids should be permitted in future test runs. The trials were designed in cooperation with special interest groups and the National Service for Special Needs Education and carried out at adult education centres across the country.

In the first part of this presentation, we will talk about some of the challenges we faced conducting the trials of test arrangements, for example the difficulty second language learners face in getting a diagnosis of dyslexia and the low number of participants who partook in the trials, as well as the results and qualitative feedback from test takers and tutors.

In the second part of the presentation, we will talk about the process that happened after the trials, where we were looking to amend the regulations of special arrangements which led us to investigating the broader question of test constructs vs. fairness, justice, and validity.

Stine Sørlie and **Hanne Lauvik** work at Skills Norway, a directory belonging to the Norwegian Ministry of Education. Skills Norway is responsible for developing the test in Norwegian for adult immigrants, a digital test that measures at Levels A1–C1. They work together to ensure that test takers with special needs have an equal opportunity to document their language skills with good quality tests. An important part of this work is to develop different test alternatives, both digital and pen-and-paper tests, to address different needs, as well as making sure that the test regulations for special needs are fair and just. They are also responsible for the tests offered to immigrant sign language users, as well as planning and running annual seminars for sign language teachers and raters on assessment of sign language skills according to the CEFR. A recent focus has been to revise and amend the arrangements for dyslexic candidates in digital reading and writing tests.

Exámenes adaptativos. Nuevos retos para la Evaluación. LC USAL esPro. Un estudio de caso

Marta García, Universidad de Salamanca

LanguageCert USal esPro es un ejemplo de cómo la Evaluación y la Tecnología van de la mano para poder afrontar los retos de la certificación lingüística de este siglo. En esta charla, mostraremos la evolución de un examen que nació en formato en papel y que ahora se ofrece por ordenador en versión adaptativa y con vigilancia remota. Hablaremos de los retos que tuvimos que superar para garantizar que la adaptación del examen mantuviera los estándares exigidos en la certificación lingüística

Marta García es profesora de español LE/L2 en Cursos Internacionales de la Universidad de Salamanca. Doctora en Filología Hispánica, Máster en Enseñanza del español y Licenciada en Filología alemana. Desarrolla su actividad principal en el área de Evaluación de Cursos Internacionales de la Universidad de Salamanca como responsable del examen LC USAL esPro.

Working towards positive washback: Aligning mediation assessment tasks to the CEFR

Caroline Shackleton, University of Granada

Adolfo Sanchez Cuadrado, University of Granada

Nathan Turner, University of Granada

It has been claimed that current language proficiency tests do not represent how languages are actually used in contemporary society (Siedhofer 2011), and so are becoming increasingly obsolete in our globalized world (McNamara and Shohamy 2016). Indeed, the publication of the CEFR Companion volume (2020) reflects just such an awareness, placing as it does even greater emphasis on the student as a social agent, one who operates within highly specific contexts. As such, it presents 'a complex vision of the situated and integrated nature of language learning and language use' (Piccardo et al 2019). The events of the past year have highlighted just how much specific language use contexts can be subject to rapid change, as educational bodies are forced to establish online responses to the challenges brought

about by the Covid-19 epidemic. Looking to the future, it seems clear that in the longer term we will witness an even stronger shift in pedagogy towards more blended, contextually-tailored learning environments. How then, should language testers respond to the difficulties posed by such an uncertain and rapidly shifting situation? Clearly, test tasks must be authentic and context-relevant if we are to make valid inferences about students' performances in real-world target language use domains, and a direct link established with real-life language functions. Here, the inclusion of mediation tasks in language assessment procedures has been hailed as an opportunity to encourage more context-specific positive washback (O'Sullivan 2018). Indeed, a mediation-orientated focus may be viewed as a common link between the many varied language use contexts we wish to evaluate, as it naturally lends itself to an integrated, scenario-based task development cycle. With these considerations in mind, this presentation looks at how intra-linguistic mediation may be incorporated into language proficiency testing in order to contribute to a more robust test design methodology, one capable of responding to the demands of the ever-changing, digital, multilingual world we live in.

Caroline Shackleton is a teacher and language testing professional presently working at the University of Granada's Modern Language Centre, where she co-ordinates the CertAcles English exam. She holds an MA in Language Testing from the Lancaster University and a PhD in Applied Linguistics from the University of Granada. She is an expert member of ACLES (Association of Higher Education Language Centers in Spain) and regularly provides training in language test development.

Adolfo Sánchez Cuadrado holds a BA in Translation and Interpreting (ES, EN, DE) from the University of Granada, an MA in Foreign Language Pedagogy from the University of Delaware and a PhD in Linguistics from the University of Granada. He is a lecturer in Linguistics at the University of Granada. He has also taught at the University of Delaware 1999–2000, at the Modern Language Centre of the University of Granada 2000-2016, and at University College London 2016-2018. His research interests lie within the fields of pedagogical translation, linguistic mediation, cross-linguistic cognitive grammar, and language teacher training. Among his publications are chapters in Methodological Developments in Teaching Spanish as a Second and Foreign Language (Cambridge Scholars Publishing, 2012), in Lingüística cognitiva y ELE (Routledge, 2019) and in Discurso y comunicación en el aula (UDIMA-EnclaveEle, forthcoming). His doctoral thesis on pedagogical translation received the 2016 Thesis Award by the International Association of Spanish as a Foreign Language (ASELE).

Nathan Turner is a teacher and test developer at the University of Granada's Modern Language Centre. He holds an MA in English Language from Lancaster University.

Using multilevel path analysis to model factors affecting collocational knowledge

Ivy Chen, Language Testing Research Centre, University of Melbourne

Collocations are essential, ubiquitous and naturally-occurring in both written and spoken English. Thus, incorrect use negatively affects fluency and marks even highly proficient learners as 'odd' or 'foreign', even if they produce grammatically correct utterances. Unfortunately, tests found in the literature suffer from drawbacks, most notably issues with unsystematic item selection. This is, in part, due to a dearth of research on factors affecting collocation difficulty. The current research thus aimed to to develop a corpus-driven receptive test of knowledge of word parts of high-frequency collocations in English, by (a) considering how the different properties of collocations (e.g., frequency effects, semantic transparency, collocation type) affect collocation difficulty and (b) using the argument-based approach to validity.

I first created a list of the 2,000 most frequent collocations based on COCA and checked against the BNC (two large corpora). Target collocations were selected from this list and consisted of compound words and four collocation types: verb-nouns, adjective-nouns, noun-prepositions, and other phrases. Participants were mostly students studying at an Australian university: 399 non-English-dominant, 148 English-dominant.

This presentation will focus on the properties of collocations that significantly predicted collocation difficulty and how their relationship can be modelled. First, I investigated the construct and found all four collocation types to be psychometrically similar to each other and to compound words. Second, multilevel path analysis showed that predictability (or its less-ideal symmetric alternative, degree of coherence) and semantic transparency significantly predicted collocation difficulty; at the other two levels, collocation type and test-taker proficiency did the same and mediated the effects of the other variables.

In sum, the systematic item selection process followed in this paper made it possible to create guidelines for item selection in future tests of collocational knowledge. Findings provide strong support for the inclusion of multiple collocation types ranging in semantic transparency and predictability.

Ivy Chen is a Research Fellow at the Language Testing Research Centre, University of Melbourne. As a former EFL/ESL teacher, her research interests cover second language acquisition and assessment. Thus, her PhD dissertation, titled A corpus-driven receptive test of collocational knowledge, combined the two areas. She used the argument-based approach to validation and modelled how properties of collocations affect item difficulty. She has tutored and lectured in applied linguistics and ESL.

Integrated reading-into-writing tasks: Do they really generate authentic and complex cognitive processes?

Alina Reid

This presentation will discuss the results of a study which investigated the cognitive processing experienced by second language English learners while engaged in multi-text integrated reading-into-writing summary tasks. Specifically, the research focused on the comparability of cognitive processing in test conditions with that experienced in real-life academic situations.

As a form of interactionist approach which targets multiliteracies instead of isolated skills, integrated reading-into-writing tasks (IRWTs) are seeing increased uptake in the field of writing assessment. Their popularity is propelled by the growing recognition of the underlying multiliteracies, multicompetencies and multilingualism that characterise real-world language use in the 21st century. IRWTs promise improved test authenticity and broader coverage of the academic domain by eliciting authentic and complex cognitive processing. As a result, many large and small-scale university tests now feature integrated skills tasks. However, despite their great potential for solving some of the most pressing issues in writing assessment, empirical evidence supporting these claims is still limited and sometimes conflicting.

This study sought to answer the call for empirical evidence to advance our understanding of integrated assessment. It used existing cognitive models of real-life

academic writing from sources in conjunction with qualitative methodology to investigate whether IRWTs generate authentic cognitive processing and complex knowledge-transformation strategies. The findings serve to substantiate the claim that IRWTs activate many of the cognitive processes relevant to the academic domain. Although several processes were under-represented, evidence of 39 out of the 45 targeted cognitive processes was found, together with five major knowledge-transformation strategies. The study uncovered insightful trends which serve to illuminate the construct of IRWTs, allowing for more confident and precise extrapolation of test results and also contributing towards empirically informed IRW task design. Implications for different testing contexts are discussed.

Alina Reid is a highly qualified ELT professional with extensive experience as teacher, teacher trainer and language test developer in a variety of educational contexts in Spain, Vietnam and the UK. She completed the Language Testing MA programme with Lancaster University in 2019, and was awarded the Geoffrey Leech Outstanding MA Student Award. As an assessment specialist, her interests lie in assessing English for Academic Purposes and authenticity in performance-based testing.

Mediating communication by plurilingual speakers in Madrid: a roadmap to assessing online and on-site social practices in post-Covid times

Antonio Ramos-Álvarez, Universidad Complutense Madrid

In this paper, we will present data analysis drawn from the language corpus of a case study in Spanish second language and culture acquisition, the informants of which make up a homogeneous experimental group: pre-university students attending Spanish lessons in Madrid (A2+ CEFR) and speakers of German (L1) and Turkish (HL), who are motivated to pursue undergraduate studies in Spain upon reaching the required entry level (B2 of the CEFR).

The study focuses on pluricultural social interactions that require the students to mediate with an array of plurilingual interlocutors in order to attain their communicative goals. Our research question is to what degree mediation provided by Spanish native speakers is more effective than mediation provided by non-native speakers of other cultures who have already participated in similar communicative events.

The research methodology includes preliminary surveys, interviews and pedagogical interventions at an early stage, followed by four participant observations carried out in planned outings around Madrid, combined with audio journals recorded by the learners so as to reflect on their own progress and self-assess their intercultural communicative competence development.

In the course of those outings, audio and video samples were recorded so as to identify the students' negative cultural perceptions that might have interfered in their communicative effectiveness. This language corpus has been assessed following two different scales adapted from the CEFR Companion Volume as regards mediating communication. A qualitative comparative analysis of the results obtained will shed light on how mediation effectiveness can be assessed accordingly, thus answering our research question.

Last but not least, we will provide some guidelines on how to collect data using (a)synchronous communication tools according to the nature of the social practice –whether virtual or physical. Given the exceptionality of the current Covid-19 times, a roadmap for the adaptability of assessment tools for both on-site and online social practices will be put forward for further discussion .

Antonio Ramos-Álvarez has a BA in English (Universidad de Salamanca), MA in Teaching Spanish (Universidad Internacional Menéndez Pelayo) and is a PhD Candidate in Spanish Language and Literature (Universidad Complutense Madrid). He has taught at university level in the USA, Turkey and Spain, and at the Cervantes Institute (CI) centres in India and Turkey. In 2015 he was appointed Spanish Programme Coordinator at the UN Secretariat. In 2016–2017 he was awarded the prize for 'The best teaching practice based on ICT' (Universidad Politécnica Madrid) and in 2018 he received the prize for 'The best teaching practice based on the use of online communication tools' (Nebrija University Madrid). In 2019, he was accredited by the Spanish Ministry of Economic Affairs and Digital Transformation as Certified Consultant in Digital Content Technologies and has since worked as EdTech consultant for several higher education institutions.

Scholarship Winning Presentations

Item performance in context: Differential item functioning between pilot and formal administration of the Norwegian Language Test

Ga Young Yoon, Norwegian Directorate of Education and Training

The Norwegian Language Test (No: Norskprøven) administered by Skills Norway is a high- stakes assessment, the results of which are used by test-takers in various ways. However, the item parameters used in multistage testing in Norskprøven are calibrated from a low-stakes situation, the pilot test. Potential item parameter shift from the pilot test to the formal test might be a concern of practitioners since it reduces the test's reliability and validity. In this study, differential item functioning (DIF) was examined between the pilot and formal reading comprehension tests in Norskprøven, using a log-likelihood ratio test method. A purification method was conducted to clean the invariant items and to improve the precision of DIF item detections. The results revealed 10 DIF items with a large effect size. A dif-ferent amount of DIF was found in different levels of ability, i.e., non-uniform DIF. DIF items also showed a tendency to vary more in item discrimination than in item difficulty. Lower discrimination parameters in the pilot test indicated more random error and might be connected to another factor, e.g., low motivation. Regarding the item features and context, i.e., item format and count of words, there was no clear evidence of being related to DIF. However, more items among the anchor items were piloted in two different levels, in contrast to the DIF items. Therefore, test administration and calibration design seem to be more related to the shift in individual item performance rather than to the item features.

Ga Young Yoon holds a Master of Science in Assessment, Measurement and Evaluation, which is cooperated by University of Oslo and Center for Education Measurement in Oslo. She obtained her degree in Bachelor of Psychology from University of Oslo, with specialisation in cultural and community psychology. Her presentation for ALTE will be about her master's thesis, where she analyses the response data from the Norwegian language test for adult immigrants, which is operated by Skills Norway. She is currently working at the Norwegian Directorate of Education and Training, where she works with developing new school exams according to the new national curricula.

YUNUS EMRE ENSTİTÜSÜ

Statistical procedures for enhancing the fairness of TYS

Burak Aydin, Ege University in Turkey

Mehmet Kaplan, Institutional Research at New Jersey Institute of Technology

The Turkish Proficiency Exam (TYS, Org. Türkçe Yeterlik Sınavı®) has been administered three times a year since 2013. The number of test-takers and test centers worldwide has been increasing because of the demand for learning the Turkish language. In 2019, 2,170 individuals took the exam in 44 centres located in 39 countries. The increasing demand has also accelerated research activities to ensure test fairness. In this presentation, the aim is to briefly mention three statistical procedures applied to the reading section of TYS. First, an application of differential item functioning (DIF) and measurement invariance with binary data is presented. Specifically, the goal is to examine and reveal possible DIF sources caused by the language family membership, namely, the Turkic family group vs. the non-Turkic family group. Second, as a potential strategy to detect cheating, attempts to locate excessive similarities in test-takers' response patterns are shortly described. Third, a comparison of different test-free measurement procedures for the reading section of TYS is presented. These procedures include the utilization of factor analyses and Item Response Theory (IRT) models for binary data to compute total scores. The former includes a traditional confirmatory factor analysis (CFA), a second-order CFA, and a bifactor model. The latter includes one-, two-, and three-parameter models in IRT along with the Rasch model. Overall, all three statistical procedures attempt to enhance the fairness of TYS and detect sources of variation that are irrelevant to the construct measured by the TYS reading section.

Burak Aydın is a faculty member in the assessment and evaluation program in the College of Education at Ege University in Turkey. He obtained a PhD degree in research methodology and a PhD minor degree in applied statistics. His research focuses on the theory and application of structural equation modeling, multilevel modeling, and propensity score analyses. He has expertise in Monte Carlo simulation studies, R programming, and complex longitudinal surveys.

Mehmet Kaplan is the Associate Director of Institutional Research at New Jersey Institute of Technology. He completed his Master's degree in Statistics and his PhD in Educational Statistics and Measurement from Rutgers University. His current research focuses on formulating and conducting statistical studies addressing issues of institutional effectiveness in higher education. In addition, he is also interested in several topics related to psychometrics such as cognitive diagnosis modeling, computerized adaptive testing, and Item Response Theory models.

PROCTOR EXAM

Online proctoring: what opportunities beyond Covid-19?

Alice Niezborala

While being around since early 2000, online proctoring experienced an incredible growth during the past year in response to the urgent need to move assessments towards a remote delivery model as a business continuity measure.

During this talk, we will be discussing insights gathered by running more than 3 million exams in 12 months and analyse the impact on ways of working and results across more than 150 partners' institutions.

More importantly, we will focus on the importance of involving all the stakeholders in the implementation process to ensure that the full potential of the technology can be leveraged to continue to improve assessment delivery and empower both candidates and test organizers through increased options and flexibility.

Finally, as interactive technologies such as video continue to evolve, we will present the opportunities to develop new hybrid forms of assessments to better engage and improve the assessment experience for learners.

Landing in edtech in 2015, **Alice Niezborala** joined ProctorExam in 2018 to continue developing an approach to localized online proctoring in Europe. As the Chief Operating Officer (COO), Alice has an active role in building the team and ways of working to support the growing operations necessary for ProctorExam development.