

Language Assessment Fit for the Future

Madrid, Spain, 26-28 April 2023

8th

International
Conference

MADRID
26-28 APR 2023



ALTE

Conference Programme



COUNCIL OF EUROPE

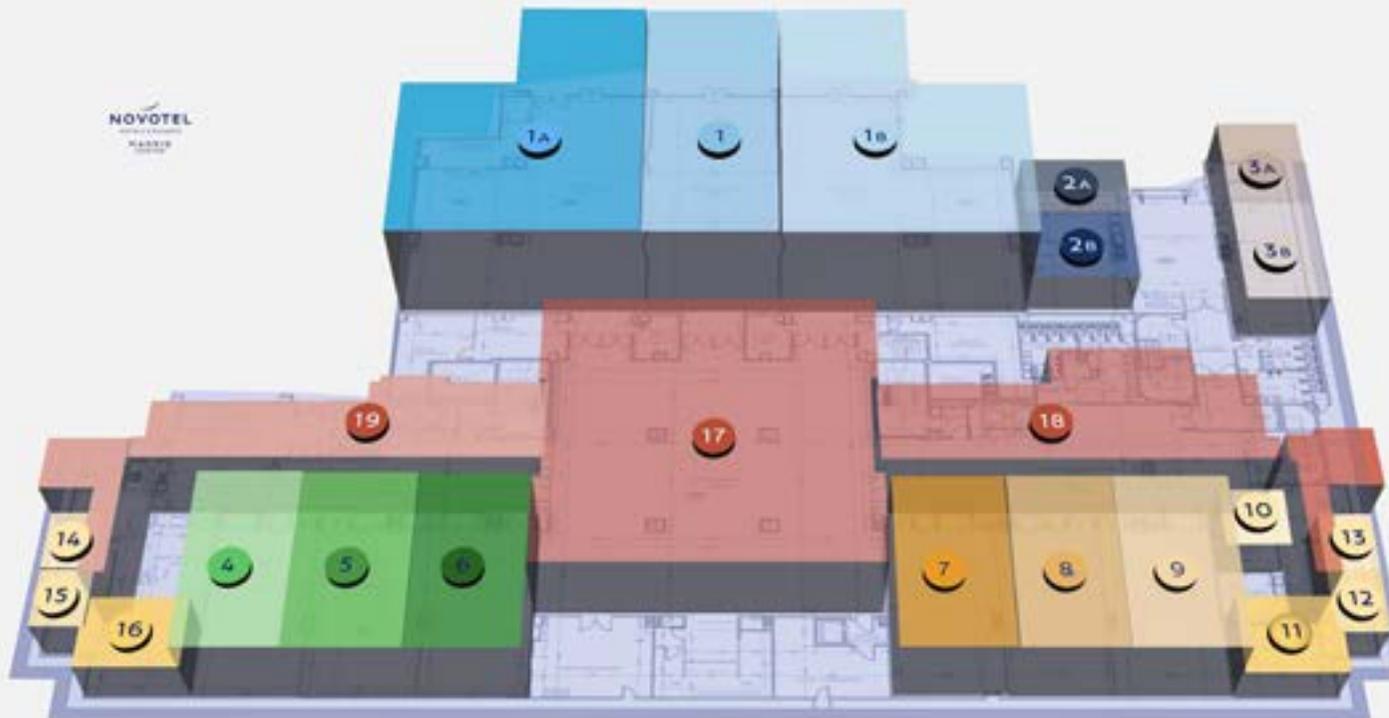


CONSEIL DE L'EUROPE

Under the auspices of the Secretary
General of the Council of Europe,
Ms Marija Pejčinović Burić

The conference rooms can be found on the Basement floor (-1) of **Novotel Madrid Centre**.

BASEMENT FLOOR PLAN



- | | | | |
|------------------------------|-----------------|-------------------|----------------------|
| 1 SALÓN CONVENCION A | 4 SALÓN VIENA | 10 MEETING ROOM 1 | 17 VESTIBULO |
| 1A SALÓN CONVENCION B | 5 SALÓN MADRID | 11 MEETING ROOM 2 | 18 VESTIBULO SALAS 1 |
| 1B SALÓN CONVENCION C | 6 SALÓN OSLO | 12 MEETING ROOM 3 | 19 VESTIBULO SALAS 2 |
| 2A SALÓN REUNIONES LONDRES A | 7 SALÓN CARACAS | 13 MEETING ROOM 4 | |
| 2B SALÓN REUNIONES LONDRES B | 8 SALÓN LA PAZ | 14 MEETING ROOM 5 | |
| 3A SALÓN REUNIONES PARÍS A | 9 SALÓN BAHÍA | 15 MEETING ROOM 6 | |
| 3B SALÓN REUNIONES PARÍS B | | 16 MEETING ROOM 7 | |

The **Plenary Room** is 1+1A+1B (Salón Convención)

Breakout Rooms are:

2A+2B	Londres
3A+3B	París
4	Viena
5	Madrid
6	Oslo
7	Caracas
8+9	La Paz+Bahía

Catering and Exhibition stands will be located in the central area marked **Vestíbulo** on the below floor plan.

The **lunch on Thursday** will be served in both of the hotel's restaurants, Roble Lima and Duque, which are located on the ground floor of the hotel.

Organisations Exhibiting:

ALTE
 Instituto Cervantes
 Yunus Emre Institute
 Goethe-Institut
 Duolingo
 Cambridge University Press & Assessment
 British Council
 telc
 GradeMaker
 Televic
 IELTS
 SIELE

Table of contents

<u>General Information</u>	4	<u>Abstracts and Bios</u>	16	<u>Social Justice Forum</u>	71
<u>Introduction</u>	5	<u>Opening Plenary</u>	16	<u>Schedule</u>	71
<u>Welcome to Madrid and to the 8th International Conference of ALTE</u>	5	<u>Keynote Speakers</u>	16	<u>Abstracts and Bios</u>	72
<u>Welcome message from Instituto Cervantes</u>	6	<u>Featured Speakers</u>	17	<u>LAMI Forum</u>	74
<u>Themes and Strands</u>	7	<u>European Commission</u>	19	<u>Schedule</u>	74
<u>Conference Schedule</u>	8	<u>Platinum Sponsor Session</u>	19	<u>Abstracts and Bios</u>	75
<u>Wednesday 26th April</u>	8	<u>Panel Sessions</u>	19		
<u>Thursday 27th April</u>	10	<u>Member Showcase Presentations</u>	22		
<u>Friday 28th April</u>	14	<u>Sign Language Presentations</u>	26		
		<u>Workshops</u>	27		
		<u>Scholarship Winners' Presentations</u>	28		
		<u>ALTE Information Sessions</u>	30		
		<u>Paper Presentations</u>	31		
		<u>Wednesday 26th April</u>	31		
		<u>Thursday 27th April</u>	37		
		<u>Friday 28th April</u>	60		
		<u>Video Short Presentations</u>	66		

General Information

Conference Times

Wednesday 26th April 2023	12:00-18:00
	Sessions begin at 14:00
Thursday 27th April 2023	08:15-17:30
	Sessions begin at 09:00
	Gala Reception at 20:00
Friday 28th April 2023	08:30-17:30
	Sessions begin at 09:00

The Gala Reception, kindly sponsored by Instituto Cervantes, will take place at **Instituto Cervantes Building**, Calle de Alcalá 49.

Information Desk

The ALTE stand and Information Desk will be located in the exhibition area and will be operational during conference opening times. Visit the Information Desk if you have any queries regarding the conference or your stay in Madrid.

Refreshments and Lunch

Complimentary tea/coffee will be served during the coffee breaks every day for registered delegates. Lunch will be provided on Thursday 27th April only.

If you stated at the time of registration that you would like a vegetarian option or have special dietary requirements, these will be catered for. If you have not informed us, please contact secretariat@alte.org immediately.

Welcome Performance

Wednesday 26th April, 18:00-18:30

Please join us in the **Salón Convención on Wednesday 26th April at 18:00** for a special performance kindly sponsored by our conference partner, Instituto Cervantes.

Gala Reception

Thursday 27th April, 20:00-23:00

The **Gala Reception** is a free event for registered delegates, kindly sponsored by our conference partner, Instituto Cervantes.

The reception will include drinks and a hot buffet dinner and take place at **Instituto Cervantes Building**, Calle de Alcalá 49.

Your delegate badge must be displayed when entering the reception.

Useful contacts

For any queries, please contact the **ALTE Information Desk** or email secretariat@alte.org

Novotel Madrid: +34 91 221 40 60

Emergency contact details (Wednesday 26th April – Friday 28th April ONLY)

Mariangela Marulli: +44 (0) 7725 781639
Sabine Edwards: +44 (0) 7984 510856

Welcome to Madrid and to the 8th International Conference of ALTE

It is wonderful to be back here in Madrid in person, especially after the disappointment of having to postpone our event in 2020. That was in the early days of Covid when we hadn't yet come to terms with the disruption and uncertainty it would cause. Much has changed in the past three years, partly due to that disruption, but also due to rapid advances in technology, such as the arrival on the scene of AI tools like ChatGPT, and the publication of the CEFR's Companion Volume with a renewed focus on concepts such as mediation and systemic alignment.

We are looking forward to reflecting on some of these developments through the overarching theme for this year's event and in grasping the opportunities for innovation created by the "new normal" after Covid. In this respect, the conference is taking place in a partly **hybrid format**, with some digital provision for those who are not able to travel on this occasion.

2020 was notable as the 30th anniversary of the first official meeting of ALTE and we had hoped to be celebrating that special occasion in collaboration with the Instituto Cervantes, one of our long-standing Members from Spain. We were grateful to Cervantes for their support in organising the planned event back in 2020 and now I would like to reiterate our thanks again on this occasion.

A big thank you to all the colleagues in Madrid for their hard work and resilience and especially for making their wonderful building available for the gala networking reception on Thursday evening.

ALTE's relationship with Spain dates back to 1989 when exploratory discussions took place between representatives from the University of Salamanca and the University of Cambridge. That encounter led to an inaugural meeting of the founder Members of the Association the following year in Barcelona.

In marking the occasion, ALTE published a booklet that traces the history from those early days and outlines the evolution of the Association, with insights and reflections from Members past and present. The booklet describes ALTE's extensive projects and collaborative activities, including nearly 70 official meetings, conferences and larger International Conferences hosted in over 40 locations across Europe; not to mention interest group meetings and courses on subjects related to language testing that ALTE has held over the past 30 years. Please take a look at it on the website www.alte.org/Our-history.

ALTE is now looking forward to a bright future over the next 30 years! In 2019, after a period of review and consultation with Members and stakeholders, ALTE completed a transformation of its legal status to become a **Charitable Incorporated Organisation (CIO)**. Not only does this enable wider participation within ALTE, but it also allows for new types of membership and other forms of collaboration. There will be opportunities at this event to discuss these with delegates.

Some things, however, are not changing. Above all, **ALTE's commitment to multilingualism and diversity remains at the centre of our mission**, and this event is a reaffirmation of that intent. ALTE continues to be the leading forum for considering these issues in the context of language learning and assessment, and so it is a great pleasure to be introducing this excellent programme with its overarching theme of '**Language Assessment Fit for the Future**' and its three central strands of **Digital, Frameworks and Diversity & Inclusion**.

Our Keynote speakers are leading voices in their own fields, and I am grateful that they agreed to set the scene and to share their views with us. I also welcome the wide array of relevant topics that were submitted as proposals for peer review and am pleased to see the important work and talented speakers that have been accepted into the programme.

In addition to the three Keynote speakers, we have ended up with three Featured Speaker Sessions, four Panel Sessions, three Workshops, presentations from the five Scholarship Winners, and five Member Showcase Presentations. We have also accepted 98 refereed paper presentations, and for the first time, we are featuring 13 "Video Shorts" which allow those who are unable to be in Madrid in person to present their research.

Plurilingual communication and translanguaging are the norm at ALTE events, whether in the conference room or during the coffee breaks. ALTE's commitment to multilingual practices and diversity is also manifested in the range of languages that are to be used during the presentations and workshops. We are delighted to welcome back members of the sign language community and two presentations relating to this area take place in the Main Auditorium on Thursday morning. Interpretation into ISL will be available for some of the main sessions and in other cases on request.

Another core element of ALTE's mission is **the fair and ethical use of language assessment**. Not only is this picked up in the strands of the conference theme again, we are delighted to host Forums on these topics this time. **The Forum on Language Assessment for Migration and Integration (LAMI)**, returns on Friday afternoon, this time on the topic of language testing and advocacy within the migration context. **The Social Justice Forum** is a new feature in the programme and will take place for the first time on Thursday afternoon. More information on both forums can be found at: www.alte.org/LAMI-Forum.

I therefore wish you a very enjoyable and thought-provoking conference, stimulating you professionally and academically, and hope you have some time to see the beautiful city of Madrid. ¡Viva Madrid!

Nick Saville, ALTE Secretary General, April 2023

Mensaje de bienvenida del Instituto Cervantes

Welcome message from Instituto Cervantes

Es un honor para mí dar la bienvenida a los participantes en la 8.ª Conferencia Internacional de ALTE. Tras más de treinta años de trayectoria, ALTE ocupa actualmente un lugar único en el ámbito de la certificación y la evaluación de lenguas en Europa. Su aportación al conocimiento en la especialidad es universalmente reconocida y los resultados de su trabajo han permitido un notable avance en el desarrollo de procedimientos y herramientas que permitan garantizar la validez y la fiabilidad de la certificación lingüística.

En esta conferencia se va a debatir sobre la adaptación de la certificación lingüística a la era digital, la implementación de marcos de referencia y el tratamiento de la inclusión y la diversidad, objetivos de primer orden en una disciplina que debe adaptarse de forma justa e inclusiva a las nuevas exigencias de la sociedad. Estoy segura de que las conclusiones que se alcancen en los debates, con el aval del conocimiento experto de los miembros de ALTE, van a contribuir al logro de estos objetivos y servirán de estímulo para abrir caminos en esta importante faceta de la investigación en lingüística aplicada.

Para el equipo profesional del Instituto Cervantes será un placer tener la ocasión de atender y acompañar a todos los participantes durante su estancia en Madrid, una ciudad siempre abierta a la innovación.

El Instituto Cervantes reafirma, con ocasión de esta nueva conferencia internacional de ALTE, su compromiso con una visión de las lenguas como herramientas privilegiadas para avanzar en el logro de un mundo basado en el entendimiento, la tolerancia, la igualdad de oportunidades y la consolidación de los valores democráticos.

Desde el Instituto Cervantes, con nuestro mayor aprecio y reconocimiento, ¡bienvenidos a Madrid!

Carmen Noguero Galilea
Secretaria General del Instituto Cervantes

It is my great honour and privilege to welcome all of you to the ALTE 8th International Conference. With over 30 years' experience, ALTE is today a well-reputed organisation in the field of language assessment in Europe. Its contribution to the advancement of testing knowledge is widely recognised, and the work of their members has helped shape the procedures and tools required to guarantee the validity and reliability of language testing.

Discussions at the conference will revolve around the adaptation of language assessment to the digital age, the implementation of reference frameworks, and the attention to diversity, first-order objectives in a discipline that must adapt to the new demands of society in a fair and inclusive manner. I firmly believe that the conclusions reached at the end of the debates, with the expert endorsement of ALTE members, will be yet another step towards the attainment of our objectives and will pave the way for further relevant research in applied linguistics.

The team of professionals at the Instituto Cervantes will gladly assist and accompany all participants during your stay in Madrid, a city always open to innovation.

Against the auspicious backdrop of this new edition of the ALTE International Conference, I would like to seize the opportunity to reiterate our commitment to language conceived as an unparalleled platform for mutual understanding, the promotion of tolerance and equal opportunities, and the strengthening of democratic principles.

Your presence here today is much appreciated. On behalf of the Instituto Cervantes, I extend a very warm welcome to everyone assembled in Madrid.

Carmen Noguero Galilea
General Secretary of the Instituto Cervantes

Themes and Strands

ALTE's mission is to set standards for language assessment, sustain diversity of languages (both spoken and signed) through our commitment to multilingualism, and maximise impact of language teaching, testing and assessment in Europe and beyond. Our eighth international conference considers how language assessment shapes, and is shaped, by wider society, in order to be fit for the future in this decade.

The theme is divided into three strands.

Fit for the Digital Age

The digital age has been with us for some years now but Covid proved a catalyst for an acceleration of all things digital which reached into language assessment just as it did in many other areas of life. Language assessment bodies needed to change quickly while still having a longer term view of what all this means for the industry and ultimately for the language learner themselves.

Papers in this section may include, among other areas:

- Technological developments in language assessment;
- Automarking;
- Auto test construction;
- Remote proctoring;
- Use of AI;
- Ethics in technology.

Diversity and Inclusion in Language Assessment

Language assessment organisations may be prone to criticisms of being too focused on protecting the interests of the culture of large languages. How can we balance the need of assessing standardised languages to prove communication ability, yet recognise the diverse nature of all individuals who use language? How can we protect and advocate those who use a so-called 'minority' language, or those who require additional support when being assessed? How can we keep language assessments ethically, morally, socially and environmentally sound?

Papers in this section may include, among other areas:

- Making assessment relevant and understood in diverse contexts;
- Lesser-used languages, dialects and variants;
- Sign languages;
- Social justice;
- Assessment for immigrants;
- Special needs and requirements;
- Issues of equality, diversity, inclusion and belonging;
- Lifelong learning;
- Sustainability.

Implementation of Frameworks

In language education, frameworks help a common understanding of what may be learnt, taught or assessed. The development and increased use of such tools, notably but not limited to the CEFR, should enable better quality education and assessment for language users. It enables description, explanation and application of constructs of language and even other social, cognitive and cultural behaviours that affect language. Nevertheless, implementation of frameworks come with challenges and these need to be overcome to have benefit for our learners and candidates.

Papers in this section may include, among other areas:

- Alignment of education and assessment to frameworks;
- Application of the CEFR Companion Volume;
- Mediation;
- Pluriculturalism and plurilingualism;
- Adaptations of the CEFR;
- Other frameworks for language assessment (e.g. CSE in China).

Conference Schedule

WEDNESDAY 26th APRIL 2023

Time	Salón Convención (1) (Live streamed) (In-person: Interpretation into Spanish/English provided; Interpretation into ISL provided on request)	Londres (2)	Paris (3)	Viena (4)
12:00-14:00	Registration			
14:00-14:10	Welcome & Housekeeping <i>Graham Seed, ALTE Secretariat Manager</i>			
14:10-14:20	Welcome <i>Carmen Noguero Galilea, General Secretary of Instituto Cervantes</i>			
14:20-14:45	Opening Plenary: Language Assessment Fit for the Future <i>Nick Saville, ALTE Secretary-General</i>			
14:45-15:30	Keynote Speaker: Back to basics to be fit for the digital age: Reflecting on advancements, opportunities and challenges of technology-enhanced language assessment [D] <i>Tineke Brunfaut</i>			
15:35-16:00	ALTE Information Session: What is ALTE and how can we help you? <i>Graham Seed and Kasia Wozniak</i>	AIDA – AI in Digital Assessment – Lessons learnt from 3 use cases of converting AI research results into usable language assessment features [D] <i>Bert Wylín and Filip Vanlerberghe</i>	Recognising stakeholder interdependency through a language assessment literacy training framework in specific purpose language assessment [F] <i>Neil Bullock</i>	Member Showcase Presentation - HACE: HACE C1: Evidencias y garantías de validez en la acreditación de la competencia comunicativa en euskera [DI] <i>Paula Elosua, Iñaki Villoslada y Enara Azkue</i>
16:00-16:30	COFFEE BREAK AND EXHIBITION			
16:30-16:55	ALTE Panel: Association of Language Testers in Europe (ALTE): promoting, maintaining, and safeguarding quality in language testing <i>Anne Gallagher, Nick Saville, Vincent Folny and Waldemar Martyniuk</i>	Early perspectives on high-stakes assessment conducted online; the complex relationship between test security and rating candidate performances in an at-home setting [D] <i>Tony Clark and Emma Bruce</i>	Post-entry language assessment in a diverse open access university system: the role of academic and general language ability [DI] <i>Jordi Heeren and Lieve De Wachter</i>	
17:00-17:25		The Magic of Machine Learning: How to Develop Tests in Multiple Languages Simultaneously and At Scale [D] <i>Sarah Goodwin</i>	Recommended validity periods for high-stakes language test scores: How are these understood and applied in HE, medical, and immigration sectors? [DI] <i>Emma Bruce, Karen Dunn and Tony Clark</i>	Cuestiones pendientes en la evaluación de la escritura creativa [F] <i>Jorge J. Sánchez Iglesias</i>
17:30-17:55	Featured Speaker Europe (interpretation into English provided): La situación actual de la evaluación certificativa de lenguas en España: una rápida (r)evolución <i>Javier Fruns y Manuela Gil-Toresano</i>			
18:00-18:30	Social event: music performance			

[D] Fit for the Digital Age; [F] Implementation of Frameworks; [DI] Diversity and Inclusion in Language Assessment

WEDNESDAY 26th APRIL 2023

Madrid (5)

Oslo (6)

Caracas (7)

La Paz + Bahía (8+9)

Time

Registration

12:00-14:00

14:00-14:10

14:10-14:20

14:20-14:45

14:45-15:30

Describing washback: teachers and students' voices in Jaén (Spain) [DI]

Victoria Peña Jaenes

Standard setting across multi-mode qualifications [D]

Lauren Miller

Language tests in controlled computer rooms against those taken online with distance invigilation: some reflections [D]

Thomas Christiansen

The Test as an Opportunity. The Case of Less Widely Tested Languages [DI]

Dina Vilcu

15:35-16:00

COFFEE BREAK AND EXHIBITION

Promoting positive washback to teaching and learning through understanding of language frameworks [F]

Margaret Malone and Howie Berman

Advantages and practicalities of a flipped classroom approach to standard setting [D]

Jane Lloyd

Automated Remote Proctoring in Language Assessment: Is the Future Already Here? [D]

Larissa Leifland Berntsson

Workshop: Considerations and challenges in setting up a retrospective DE&I review of a high-stakes item bank [DI]

Kasia Wojciechowska and Lottie Harrison

16:00-16:30

16:30-16:55

Assessment of existential competence of learners in language education: Issues, frameworks, modelling and standards of performance [F]

Denise Lussier, Ph. D. and Isabelle Lussier, M. A.

Killing a flock of standard-setting judgements with one digital stone [D]

Anniken Telnes Iversen, Rønnaug Katharina Totland and Torbjørn Ohlseth

Online Proctoring of Language Examinations: A Survey of Candidates' Attitudes and Perceptions. Languagecert online ESOL exams and USAL esPro test [D]

Marta García, Yiannis Papargyris and Leda Lampropoulou

17:00-17:25

17:30-17:55

18:00-18:30

18:30-19:30

Sponsors' Reception
(Invited Sponsors only)

[D] Fit for the Digital Age; [F] Implementation of Frameworks; [DI] Diversity and Inclusion in Language Assessment

THURSDAY 27th APRIL 2023

Time	Salón Convención (1) (Live streamed) (In-person: Interpretation into Spanish/English provided; Interpretation into ISL provided on request)	Londres (2)	Paris (3)	Viena (4)
8:15-9:00	Registration for new arrivals			
9:00-9:45	Keynote Speaker: Reframing assessment literacy: putting diversity and inclusion firmly in the picture [DI] <i>Lynda Taylor</i>			
9:50-10:15	Sign Language Presentation: Assessing receptive skills development in deaf children who use Swiss German Sign Language as their primary language [DI] <i>Tobias Haug</i>	Panel Session: Implementing mediation: from theory to practice [F] <i>Joaquín Manuel Cruz Trapero, Adolfo Sánchez Cuadrado, Olga Naranjo Travieso and María Luz Castro Pérez</i>	Language tests for Citizenship: A review [DI] <i>Miguel Fernández Álvarez</i>	Member Showcase Presentation - Cambridge University Press & Assessment: How can we make language assessment fit for the future? Current trends and future opportunities [D] <i>Angeliki Salamoura</i>
10:20-10:45	Sign Language Presentation: Common European Framework of Reference and Czech Sign Language, project 2019-2022 [DI] <i>Vladimir Simon and Denisa Lachmanova</i>		Citizenship tests as a means of inclusion. How far have we gone till now? [DI] <i>Thomas Rousoulioti and Maria Iakovou</i>	Expanding frameworks: New developments in vocational language testing [F] <i>Magdalena Igiel</i>
10:45-11:10	COFFEE BREAK AND EXHIBITION			
11:10-11:35	Panel Session (interpretation into English provided): Acreditación de la competencia comunicativa en lenguas minorizadas y contextos de diversidad lingüística [DI] <i>Paula Elosua, Josep Ma. Montcada, Cristina Pérez-Guillot y Iñaki Villoslada</i>	Member Showcase Presentation - CNAVT: Reporting test scores in a more insightful way [DI] <i>Manon Willems and Goedele Vandommele</i>	Embedded meaning in Canada's English-language proficiency requirements for immigration, asylum and resettlement, and citizenship [DI] <i>Melissa McLeod</i>	Issues in automated scoring of language performance tests: Methodologies, practices, and ethics [D] <i>Jing Xu</i>
11:40-12:05		In-class dynamic assessment and curriculum design framework based on needs analysis for a volunteer-led adult migrant literacy program [DI] <i>Haeng-A Kim</i>	A sociolinguistic profile of the migrants applying for Portuguese citizenship [DI] <i>Nélia Alexandre, Paula Simões and Ana Espírito Santo</i>	Are Humans Redundant? Automated Scoring in Learning and Assessment [D] <i>Dr. Arum Perwitasari</i>
12:10-12:35	Platinum Sponsor Session - Yunus Emre Institute: Language Teaching and Language Assessment Activities of Yunus Emre Institute <i>Dr. Cumhur Ersin Adigüzel</i>	Reframing language learning: Introducing the Sustainability Framework for ELT [DI] <i>Helen Allen</i>	Implementing CAF measures and the scales of Functional Adequacy in the L2 Italian certificate exam Cert.it: Implications for the assessment of B1 candidates applying for citizenship [F] <i>Diego Cortés Velásquez and Elena Nuzzo</i>	Automated Scoring of Spelling Mistakes in Short Answers [D] <i>Leska Schwarz and Ronja Laarmann-Quante</i>
12:40-13:05	Featured Speaker Latin America: An initial exploration of current language policies, national and local tests in Latin America <i>Nayibe Rosado Mendinueta</i>			
13:05-14:30	LUNCH Served in both restaurants, 'Roble Lima' and 'Duque'	Video Shorts on loop: Implementing inclusive practices in the assessment of language modules in higher education [DI] <i>Marion Coderch</i> The Impact of Multilingualism on Language Learning: The Case of A2 and B2 Students [F] <i>Olga Vladislavovna Bukach, Elzara Vasilovna Gafiyatova and Marina Ivanovna Sonyshkina</i>	Video Shorts on loop: What's in a name? Examining the construct of name writing in assessing emerging writing of multilingual kindergartners [DI] <i>Gordon Blaine West and Jeanne Beck</i> Adapting the MLAT-ES into Catalan: when young learners' age is not just any other variable [DI] <i>María del Mar Suárez</i>	Video Shorts on loop: CSE-based Dynamic Assessment for English Reading in Rural Northwest China [DI] <i>Huilin Lyu</i> Embrace the Future of Minority Language Testing: Insights from Zhuang Language Proficiency Test in China [DI] <i>Andy Jiahao Liu</i>

[D] Fit for the Digital Age; [F] Implementation of Frameworks; [DI] Diversity and Inclusion in Language Assessment

THURSDAY 27th APRIL 2023

Madrid (5)	Oslo (6)	Caracas (7)	La Paz + Bahía (8+9)	Time
Registration for new arrivals				8:15-9:00 9:00-9:45
<p>Member Showcase Presentation - telc gmbH: Standardisation On-screen: Ensuring Exam Equity Across 24 European Languages [D]</p> <p><i>R. Sean McDonald</i></p>	<p>Application of the CEFR Companion Volume to a cognitive view on pragmatic competence [F]</p> <p><i>Cristina Heras Ramirez</i></p>	<p>Improving proficiency descriptors in the 21st century: Incorporating diversity, equity and inclusion [DI]</p> <p><i>Margaret Malone and Howie Berman</i></p>	<p>Investigating potential bias in testing migrants' language proficiency in Switzerland [DI]</p> <p><i>Hrisztalina Hrisztova-Gotthardt</i></p>	9:50-10:15
<p>CEFR-J x 28 multilingual can do test development: an interim report [F]</p> <p><i>Yukio Tono and Masashi Negishi</i></p>	<p>A Descriptive Overview of Pretesting Assessment within the CEFRL [F]</p> <p><i>Matthew James and Maria Isabel Requena de Lamo</i></p>	<p>A Multiple Group IRT Approach to Address Diversity and Inclusion in a Reading Assessment [DI]</p> <p><i>Ismail Cuhadar and Burcu Parlak</i></p>	<p>Language tests for migrants: How to deal with stakeholders' expectations? [DI]</p> <p><i>Sibylle Plassmann</i></p>	10:20-10:45
COFFEE BREAK AND EXHIBITION				10:45-11:10
<p>Member Showcase Presentation - Charles University Prague: Realignment of the Czech Language Certificate Exam to the CEFR: Processes, issues and implications [F]</p> <p><i>Martina Hulešová</i></p>	<p>Connecting an Academic English Integrated Writing Task to the CEFR [F]</p> <p><i>Mohammed Sameer</i></p>	<p>Evaluating an accommodations policy for candidates in a large-scale English proficiency test [DI]</p> <p><i>Johanna Motteram and Richard Spiby</i></p>	<p>ALTE Information Session: Raising Standards in Assessment Literacy - the ALTE-Equals Course</p> <p><i>Elaine Boyd</i></p>	11:10-11:35
<p>What is the future of plurilingual language assessment for 'monolingual' testing organisations? [DI]</p> <p><i>Graham Seed</i></p>	<p>Aligning a multimodal integrated speaking assessment task to the Common European Framework of Reference [F]</p> <p><i>Kim Anne Barchi and Mariana Jo Bisset</i></p>	<p>An Educational Action Plan to Improve Assistance to Autistic Spectrum Disorder (ASD) Test-Takers in Written Large-Scale Exams [DI]</p> <p><i>Anarcsa de Freitas Nascimento</i></p>	<p>ALTE Information Session: ALTE Q-Mark: Practicalities</p> <p><i>Helen Garside-Hornby</i></p>	11:40-12:05
<p>Language assessment across a hundred languages - how to ensure test quality and valid results? [DI]</p> <p><i>Mari Nielsen Vaage and Marius Lian</i></p>	<p>Alignement d'un test d'évaluation de français langue étrangère sur des cadres de référence distincts [F]</p> <p><i>Dominique Casanova et Emilie Grosset</i></p>	<p>Adapting online language activities and assessment for functionally diverse learners [DI]</p> <p><i>Jackie Robbins</i></p>	<p>ALTE Information Session: ALTE Services Unit drop-in</p> <p><i>Kasia Wozniak</i></p>	12:10-12:35
				12:40-13:05
<p>Video Shorts on loop:</p> <p>Using Dynamic Assessment of Writing to Promote Technology-Enhanced Learning in Higher Education [D]</p> <p><i>Eleni Meletiadou</i></p> <p>Towards AI-Assisted Essay Marking: Developing and Evaluating Automated Writing Scoring Models for a General English Proficiency Test [D]</p> <p><i>Michelle Chen</i></p>	<p>Video Shorts on loop:</p> <p>Formative Assessment and Student-Centred Learning: a match made in HEaven [F]</p> <p><i>Sebastian Koźbial</i></p> <p>Inclusive Formative Assessment Practices (IFAP) in Higher Education: Promoting Education for Social Justice [D]</p> <p><i>Eleni Meletiadou</i></p>	<p>Video Shorts on loop:</p> <p>Modelling information-based academic writing: A domain analysis focusing on the knowledge dimension [D]</p> <p><i>Yu Chengyuan</i></p> <p>Diagnosis and Mediation of Argumentative Writing Ability in L2 Concept-Based Debate Instruction [F]</p> <p><i>Hua Fu</i></p>	<p>Video Shorts on loop:</p> <p>Cross-Country Comparison of English Speaking Ability with PROGOS Test [D]</p> <p><i>Masuyo Ando and Yukio Tono</i></p>	13:05-14:30

[D] Fit for the Digital Age; [F] Implementation of Frameworks; [DI] Diversity and Inclusion in Language Assessment

THURSDAY 27th APRIL 2023

Time	Salón Convención (1) (Live streamed) (In-person: Interpretation into Spanish/English provided; Interpretation into ISL provided on request)	Londres (2)	París (3)	Viena (4)	
SOCIAL JUSTICE FORUM					
15:30-14:55	14:30	Welcome & Introduction: Social justice - barriers and opportunities. <i>Cecilie Hamnes Carlsen</i>	Member Showcase Presentation - Foreign Language Centre, University of Pecs: Does the mode of delivery influence test takers' performance? A comparative analysis [D] <i>Róbert Márcz and Réka Werner</i>	The use of Global Englishes in Listening assessment: A review of the evidence and future directions of research [D] <i>Amy Devine</i>	La formación y actualización en línea de los calificadoros del DELE [D] <i>Eva María García García</i>
	14:40	Mandatory integration requirements for foreigners in the EU: what role for human rights? <i>Zvezda Vankova</i>			
15:00-15:25	15:00	Language assessment in search of social justice: Is it a chimera? <i>Antony John Kunnan</i>	The Comparability of Computer-Based and Paper-Based Writing Tests: A Case Study [D] <i>Balázs Csizmadia</i>	Balancing the need for native and non-native speakers in ELF listening tasks: to what extent do accents affect comprehension? [D]	El diseño interactivo del examen del SIELE [D] <i>Aurora Duque de la Torre</i>
	15:20	Dialogue and engagement between policymakers and language test developers. <i>Laura Schildt</i>		<i>Anna Maria De Bartolo and Jean Marguerite Jimenez</i>	
15:30-15:55	15:40	Citizenship test exemptions around Spanish and Catalan. <i>Kamran Khan</i>	Multidimensional analysis of syntactic complexity of L2 students' exam writing at the B2+ level [F] <i>Karlygash Adamova</i>	A pilot material for a fair and accessible A2 listening test for adult immigrants with diverse educational backgrounds [D] <i>Elina Stordell</i>	La evaluación de tareas de revisión en entornos multilingües [F] <i>Jorge J. Sánchez Iglesias</i>
COFFEE BREAK AND EXHIBITION					
16:30-16:55	16:30	Plurilingual Profiles and Multilingual Needs of Adult Refugees and Migrants in Greece and Italy. <i>Anna Mouti</i>	Using Screencast Feedback to Improve Undergraduate Students' Speaking Skills in English [D] <i>Yéssica Aguilera Fuentes</i>	Practical model of intensive reading assessment with ReadTheory in the Reading Club [D] <i>Salwa Said Khalfan Al Anquodi</i>	El Manual oficial del SIELE y su integración con la Extensión digital del examen [D] <i>Juan Miguel Prieto Hernández</i>
	16:50	Language testing and language rights. <i>Dina Vilcu</i>			
17:00-17:25	17:10	Discussion. <i>Nick Saville</i>	Assessing L2 Speaking Performance Asynchronously and Online: On Navigating the Fields of Tension between Authenticity, Validity, Reliability and Practicality [D]	Interpret and integrate ideas and information in PIRLS [D] <i>Lena Edander</i>	Marcadores y certificación lingüística [D] <i>Marta García</i>
	17:20	Round-off. <i>Cecilie Hamnes Carlsen</i>	<i>Olivia Rützi-Joy</i>		
20:00-23:00	GALA NETWORKING RECEPTION Instituto Cervantes Building, Calle de Alcalá 49. Sponsored by Instituto Cervantes				

[D] Fit for the Digital Age; [F] Implementation of Frameworks; [DI] Diversity and Inclusion in Language Assessment

THURSDAY 27th APRIL 2023

Madrid (5)	Oslo (6)	Caracas (7)	La Paz + Bahía (8+9)	Time
<p>Calibration of a “Golden corpus” to assess the effectiveness and validity of an algorithmic solution (Artificial intelligence) for assistance of writing assessment: findings, limits, opportunitiess [D]</p> <p><i>Vincent Folny</i></p>	<p>Putting the CEFR at the centre of classroom assessment in a global teaching program [F]</p> <p><i>Aidan Holland, Carolyn Westbrook and Johnathan Cruise</i></p>	<p>Developing standardized tests for Arabic for the final exams in secondary education: experiences in the Netherlands [DI]</p> <p><i>Anneke de Graaf</i></p>	<p>Assessment in the Early Years: Mapping Concepts and Practices in four Brazilian states [DI]</p> <p><i>Juliana Reichert Assunção Tonelli</i></p>	14:30-14:55
<p>Considering inter-test relationships in high-stakes admissions testing: The case of English [D]</p> <p><i>Ramsey Cardwell and Steven Nydick</i></p>	<p>Using the CEFR descriptors as Performance Level Descriptors (PLDs): Opportunities, challenges, and insights [F]</p> <p><i>Paraskevi (Voula) Kanistra</i></p>	<p>Effects of examinee L1, assessment criteria and L1 recognition in Finnish National Certificates of Language Proficiency (NCLP) [DI]</p> <p><i>Tuija Hirvelä and Reeta Neittaanmäki</i></p>	<p>The development and reception of a digital task-based exam for young language learners [D]</p> <p><i>Anne van Asseldonk and Lies Strobbe</i></p>	15:00-15:25
<p>Piloting Automatic Computerized Testing to Evaluate the Language Abilities of Pre-service World Language Teachers [D]</p> <p><i>Michael Vrooman</i></p>	<p>A flexible framework: matching student assessments to the CEFR descriptors in a hybrid context [F]</p> <p><i>Steve Issitt</i></p>	<p>Using multi-level tests in benchmarking projects in Iberia [D]</p> <p><i>David Bradshaw and Victoria Peña Jaenes</i></p>	<p>Implementing a communicative classroom-based assessment program in Dutch lower-form secondary education: teacher and student perceptions [F]</p> <p><i>Charline Rouffet</i></p>	15:30-15:55
COFFEE BREAK AND EXHIBITION				
<p>Online language testing: investigating the candidates’ attitudes and reactions [D]</p> <p><i>Letizia Cinganotto</i></p>	<p>Mapping the SMEEA Gaokao tests to the CEFR [F]</p> <p><i>Jane Lloyd</i></p>	<p>Italian language testing regime: alternative perspectives [F]</p> <p><i>Paola Masillo, Giulia Peri and Sabrina Machetti</i></p>	<p>Teaching the teachers: Designing digital assessment for language teachers which both evaluates and educates [D]</p> <p><i>Nicole Busby and Anja Angelsen</i></p>	15:55-16:30 16:30-16:55
<p>“No-one gets left behind”- Getting stakeholders ready for the digital age [D]</p> <p><i>Isabel Knöb</i></p>	<p>The implementation of the CEFR in Sweden: the relation to the national curriculum standards for languages [F]</p> <p><i>Maria Håkansson Ramberg</i></p>	<p>The ITACE for Lecturers: a locally developed C1 test for lecturers in Flemish higher education [F]</p> <p><i>Frank van Splunder, Catherine Verguts and Tom De Moor</i></p>	<p>Language Background and its Effect on Performance on a Digital Age Test [D]</p> <p><i>Nathan T. Carr</i></p>	17:00-17:25
GALA NETWORKING RECEPTION Instituto Cervantes Building, Calle de Alcalá 49. Sponsored by Instituto Cervantes				
20:00-23:00				

[D] Fit for the Digital Age; [F] Implementation of Frameworks; [DI] Diversity and Inclusion in Language Assessment

FRIDAY 28th APRIL 2023

Time	Salón Convención (1) (Live streamed) (In-person: Interpretation into Spanish/English provided; Interpretation into ISL provided on request)	Londres (2)	París (3)	Viena (4)
9:00-9:45	<p>Keynote Speaker: Aligning language education to the CEFR: Whys, whats and hows [F]</p> <p><i>Neus Figueras</i></p>			
9:50-10:15	<p>Panel Presentation: PISA 2025 Foreign Language Assessment Framework: What matters for successful language learning and how to assess it [F]</p> <p><i>Nick Saville, Angeliki Salamoura, Catalina Covacevich, Martin Robinson and Hanan Khalifa</i></p>	<p>Scholarship Winner's Presentation: Man and Machine: Match made in heaven? (Status: "It's complicated...") – Exploring Dutch Automated Essay Scoring for young learners' written output [D]</p> <p><i>Geert De Meyer</i></p>	<p>Implementation of mediation in Catalan EOI certification exams [F]</p> <p><i>Montserrat Montagut Montagut and Josep Maria Montcada Escubairó</i></p>	<p>El efecto de los exámenes de acreditación desde el punto de vista de profesores y alumnos en Jaén (España) [DI]</p> <p><i>Victoria Peña Jaenes</i></p>
10:20-10:45		<p>Scholarship Winner's Presentation: Exploring anti-plagiarism tool effects in the assessment of academic reading-into-writing [D]</p> <p><i>Valeriia Koval</i></p>	<p>From Mediation to Knowledge Transformation: expanding the construct of the reading into writing task [F]</p> <p><i>Alina Reid</i></p>	<p>SIELE remoto: el desafío de la evaluación a distancia [D]</p> <p><i>Sandra Cruz</i></p>
COFFEE BREAK AND EXHIBITION				
10:45-11:10				
11:10-11:35	<p>European Commission Presentation: European Union policy on language learning and key data on teaching languages at school in Europe [F]</p> <p><i>Anna Solé Mena</i></p>	<p>Scholarship Winner's Presentation: Bias is everywhere? An investigation into DIF at the item, rater and task level [DI]</p> <p><i>Christine Troussart Van Bulck</i></p>	<p>Lessons from implementing mediation at the European universities based on the MiLLaT project [F]</p> <p><i>Sylwia Kossakowska-Pisarek and Malgorzata Swierk</i></p>	<p>DELE C1 y C2: prevalencia y puntuaciones de candidatos disléxicos [DI]</p> <p><i>Antonia Liberal Trinidad</i></p>
11:40-12:05	<p>Overall Scholarship Winner's Presentation: Spoken Dialog System-Delivered Paired Discussion Tasks for Interactional Competence Assessment [D]</p> <p><i>Nazlinur Gokturk</i></p>	<p>Scholarship Winner's Presentation: Including Second Language Varieties of English on High-Stakes International Tests of English Proficiency [DI]</p> <p><i>Yi Tan</i></p>	<p>Defining Alternative Constructs of Multilingual Assessment in Higher Education: The case of English in contact with other languages in mainland U.S. and Puerto Rico [DI]</p> <p><i>Eva Rodriguez-Gonzalez</i></p>	<p>Consideraciones en torno a la Evaluación en el Aula de las Lenguas desde un Marco de Justicia [F]</p> <p><i>Edgar Picón Jácome</i></p>
12:10-12:35	<p>Featured Speaker East Asia: Assessment Design for Language Teaching in Vietnam</p> <p><i>Huu Nguyen</i></p>			
12:40-13:00	<p>Scholarship Awards and Closure</p> <p><i>Nick Saville and Graham Seed</i></p>			
13:00-14:00	EXHIBITION and Registration for new arrivals for LAMI Forum			

[D] Fit for the Digital Age; [F] Implementation of Frameworks; [DI] Diversity and Inclusion in Language Assessment

FRIDAY 28th APRIL 2023

Madrid (5)	Oslo (6)	Caracas (7)	La Paz + Bahía (8+9)	Time
				9:00-9:45
<p>Overcoming Challenges Aligning Language Assessments to Standards [F] <i>Fabiana MacMillan, David MacGregor and Jason Kemp</i></p>	<p>Member Showcase Presentation - Le Français des Affaires: Combine strengths to provide high-quality digitized language tests and examinations [D] <i>Dominique Casanova and Bert Wylín</i></p>	Reserved for repeat presentations	<p>Workshop: Bigger Bang for Your Assessment Buck: How Language Counselling Enhances the Relevance of Language Assessment for All Language Learners [DI] <i>Erin McDonald and Olga Sarycheva</i></p>	9:50-10:15
<p>An online flipped classroom approach to standard setting: What factors support consistent decision making in online contexts? [D] <i>Jane Lloyd</i></p>	<p>The English component of the University Entrance Exam in Madrid under Review. Questions, Challenges and Future Directions [F] <i>Miguel Fernández Álvarez</i></p>	Reserved for repeat presentations		10:20-10:45
COFFEE BREAK AND EXHIBITION				
				10:45-11:10
<p>Calibration of new rating scales of DELF-DALF: a worldwide collaborative project [F] <i>Vincent Folny and Marie Gouelleu</i></p>	<p>Member Showcase Presentation - TestDaF-Institut: New test format – new research agenda: An overview of the technology-related research at g.a.s.t. [D] <i>Prof. Dr. Anastasia Drackert</i></p>	<p>Member Showcase Presentation - CVCL Università per Stranieri Perugia: The CELI corpus: a new tool to analyse written L2 Italian and to inform language testing and assessment [D] <i>Fabio Zanda and Danilo Rini</i></p>	<p>Workshop: Assessing lesser used languages: Papiamentu and Dutch as a Foreign Language [DI] <i>Judith Richters and Janneke Helsloot</i></p>	11:10-11:35
<p>Comparative judgement of L2 writing in learner corpora: does a narrow proficiency range reduce test reliability? [D] <i>Peter Thwaites</i></p>	<p>A comparative study of test-takers' perceptions of paper-based and online language examinations [D] <i>Róbert Márcz</i></p>	<p>Validation of a High-Stakes Test: GA IESOL Multiple-Choice Units [F] <i>Snezana Mitrovic, Emanuela Botta and Giusi Castellana</i></p>		11:40-12:05
				12:10-12:35
				12:40-13:00
EXHIBITION and Registration for new arrivals for LAMI Forum				
				13:00-14:00

[D] Fit for the Digital Age; [F] Implementation of Frameworks; [DI] Diversity and Inclusion in Language Assessment

Abstracts and Bios

Opening Plenary

Language Assessment Fit for the Future

Nick Saville, ALTE Secretary-General

WED-26th

The opening talk to the conference will focus on the overall conference theme, considering how language assessment shapes, and is shaped, by wider society, in order to be fit for the future in this decade. It sets the context for the plenary and featured speakers, as well as the other presentations, workshops and forums taking place during the conference. For language assessment to be fit for the future, it needs to have a clear understanding and ambition of theory, practice and policy in several areas, including those taken as strands within this conference: being fit for the digital age, diversity and inclusion and language assessment, and the implementation of relevant frameworks.

Dr Nick Saville is Director of Thought Leadership, English for Cambridge University Press & Assessment, and is Secretary-General of the Association of Language Testers in Europe (ALTE). His research and policy interests include: English language learning and plurilingualism; Learning Oriented Assessment (LOA) and the integration of learning with assessment (LA); EdTech combined with EdAI for the automation of language assessment; language assessment literacy; and ethical frameworks in language education. He co-edits the Studies in Language Testing series (CUP).

Keynote Speakers

Back to basics to be fit for the digital age: Reflecting on advancements, opportunities and challenges of technology-enhanced language assessment [D]

Tineke Brunfaut

WED-26th

Some of the most fundamental questions in language testing and assessment are: What are we testing? How are we testing? Why are we testing? Despite a solid and sophisticated body of research and research-informed language assessment practice, these questions and their answers remain at the heart of everything we do in the field. In this talk, I will discuss and illustrate advancements and opportunities of technology-enhanced language assessment guided by these three basic questions. This includes the role of technology in the changing nature of language use domains, in ways to describe language use domains, and in evaluating domain coverage by language tests and assessments (What?); technological innovations in administrative procedures of test development and rating, item generation and test compilation, test delivery, test security, and automated scoring and feedback generation (How?); and ethical uses of technology for language testing and assessment (Why?). At the same time, I will highlight risks of technology-driven rather than technology-enhanced practices, and discuss requirements for meaningful and justifiable uses of technology to inform critical issues and principles in the field.

Tineke Brunfaut is Professor of Linguistics and English Language at Lancaster University, UK, where she specializes in language testing and second language listening and reading research. Her work explores topics such as cognitive and affective factors in language testing, methodological innovations in language testing, and issues in language test development. Her research has been published

in journals such as Language Testing, Language Assessment Quarterly, Assessing Writing, Assessment in Education, and Applied Linguistics. With Paula Winke, she edited The Routledge Handbook of Second Language Acquisition and Language Testing (2021). Tineke is a recipient of the ILTA Best Article Award, the e-Assessment Award for Best Research, and the TOEFL Outstanding Young Scholar Award. At Lancaster, she supervises doctoral research in language testing, and teaches on the Language Testing and TESOL masters programmes. Tineke regularly conducts language test development and consultancy work for professional and educational bodies around the world.

Reframing assessment literacy: putting diversity and inclusion firmly in the picture [DI]

Lynda Taylor

THU-27th

Language testing has long been a locus for investigating the social context, consequences and power of testing and assessment. During the 1990s, exploration of language test washback and impact led to growing interest in 'assessment literacy', a concept defined broadly as the skills, knowledge and abilities that different test stakeholders (e.g. teachers, curriculum designers, policy-makers) need to carry out their assessment-related roles and responsibilities. More recently there has been growing awareness that some stakeholders in the world of language testing and assessment may not be adequately acknowledged or represented in our theory and practice. They include stakeholders working with so-called 'minority' languages, i.e. languages that are less widely used or spoken in society, or which are taught, learned and tested with relatively small population cohorts. They also include stakeholders involved in assessment accommodations for those who have special needs or requirements due to life circumstances or to a

temporary or permanent disability. This plenary talk considers how the concept of assessment literacy is continuing to evolve so as to take greater account of how specific features of testing context and purpose shape our theory and practice, and particularly the nature of communication between different stakeholder constituencies. We shall explore how assessment literacy can be reframed to give greater prominence to matters of diversity and inclusion as part of our commitment to professional ethics and social justice.

Lynda Taylor is Visiting Professor at the Centre for Research in English Language Learning and Assessment (CRELLA) at the University of Bedfordshire, UK. She holds a PhD and MPhil in Applied Linguistics from the University of Cambridge and she has worked for over 35 years in the field of language teaching, learning and assessment, particularly with IELTS and the full range of Cambridge Assessment English qualifications. Her research interests include speaking and writing assessment, testing accommodations for language learners with special needs and the development of language assessment literacy. For several years she was Assistant Research Director with Cambridge Assessment English and she has advised on test development and validation projects around the world, including the Baltic States, mainland China, Taiwan and Japan. She has presented invited plenaries, research papers and workshops internationally, published extensively in academic journals and authored or edited many of the volumes in CUP's Studies in Language Testing (SiLT) series, for which she is also series co-editor. She is an Individual Expert Member for ALTE and is currently President of the UK Association for Language Testing and Assessment (UKALTA) (2020-2022). In 2022 she was awarded Fellowship of the UK Academy of Social Sciences.

Aligning language education to the CEFR: Whys, whats and hows [F]

Neus Figueras

FRI-28th

The publication of the CEFR-CV (2020) has resulted in a renewed interest in the principles that underlie the CEFR's descriptive scheme, from which level labels and descriptors (often referred as the vertical dimension in the CEFR) have tended to divert attention. The contents and recommendations in the CEFR, what they mean, what they imply, how they can be implemented, and why they are key in language education are currently being revisited and theorized, fifty years after John Trim and David Wilkins sketched

out the action-oriented approach that became the basis of The Threshold Level' published in 1980. This session will address why it is necessary to 'go back to' and to 'go deeper into' a document which was published 20 years ago and has been widely influential but is seemingly still not fully understood. It will outline the issues raised re. the use(s) and implementation of the CEFR worldwide and in Spain in particular, by professionals working in different fields: curriculum developers, textbook writers and publishers, assessment specialists, teacher trainers or practitioners. The Handbook for aligning education to the CEFR will be discussed in this context, as it aims at contributing to a fuller understanding of the CEFR, and at facilitating the use of the document, by outlining the processes to follow when aligning curricula, materials and assessments to the CEFR and by providing guidelines to complete them. Published in April 2022, the Handbook is currently being put to use in alignment projects by different stakeholders and it is hoped that their views and experiences can be incorporated in a future edition.

Dr Neus Figueras been involved in a number of international research and development projects related to assessment at different education levels and has given courses and presented in universities in Spain, in various European countries, in Asia and in the USA. She has also carried out consultancy work in assessment and in curriculum design. She has collaborated regularly with the Council of Europe in the uses and the dissemination of the CEFR in language testing and assessment and is one of the authors of the Manual for Relating examinations to the CEFR (Council of Europe, 2009). She has recently co-edited with David Little Reflecting on the CEFR and its Companion Volume, published in 2022 by Multilingual Matters and contributed to the publication of Aligning Language Education with the CEFR: A Handbook, jointly published by the British Council, UKALTA, EALTA and ALTE, publicly available online. She is a founding member of EALTA (European Association for Language Testing and Assessment, www.ealta.eu.org), and is currently the coordinator of the EALTA CEFR SIG. She is a member of the organizing committee of GIELE (Grupo de interés en evaluación de lenguas en España, <http://giele.webs.upv.es/>).

Featured Speakers

Europe: La situación actual de la evaluación certificativa de lenguas en España: una rápida (r)evolución

Javier Fruns y Manuela Gil-Toresano

WED-26th

Esta ponencia proporciona una visión panorámica de la situación actual de la evaluación certificativa de lenguas en España, prestando especial atención al trabajo realizado en los últimos años por universidades, escuelas oficiales de idiomas y el Instituto Cervantes. En ella se analiza cómo estas instituciones han afrontado los actuales desafíos de la certificación de lenguas a través de actividades formativas que han tenido como consecuencia un mayor nivel de alfabetización en la disciplina y la puesta en marcha de proyectos cada vez más ambiciosos. A lo largo de la presentación observaremos la respuesta a retos como la incorporación de la mediación en las pruebas evaluativas, la integración de las tecnologías, el desarrollo de pruebas con fines específicos o el cumplimiento de estándares de validez y calidad. Ejemplos paradigmáticos de las acciones de formación y colaboración tanto a nivel individual como institucional son, entre otros, el Grupo de Interés en la Evaluación de Lenguas en España (GIELE), en el que los miembros han aunado esfuerzos para mejorar sus prácticas y compartir conocimientos, herramientas y estrategias; y el Sistema Internacional de Certificación del Español como Lengua Extranjera (SICELE), asociación en la que entidades educativas de España colaboran con instituciones de numerosos países hispanohablantes.

This presentation provides a panoramic view of the current situation of language testing in Spain, paying special attention to the work carried out in recent years by universities, official language schools and the Cervantes Institute. It analyzes how these institutions have faced the current challenges of language testing through training activities that have resulted in a higher level of literacy in the discipline and the implementation of increasingly ambitious projects. Throughout the presentation we will look at the response to challenges such as the incorporation of mediation in proficiency tests, the integration of technologies, the development of tests for specific purposes or the compliance with validity and quality

standards. Paradigmatic examples of training and collaboration actions both at an individual and institutional level are, among others, the Group of Interest in the Evaluation of Languages in Spain (GIELE), in which members have joined forces to improve their practices and share knowledge, tools and strategies; and the International Certification System for Spanish as a Foreign Language (SICELE), an association in which educational entities from Spain collaborate with institutions from numerous Spanish-speaking countries.

Manuela Gil-Toresano es profesora del Instituto Cervantes y formadora de profesores. Actualmente trabaja en la Unidad de Certificación Lingüística del Instituto Cervantes en Madrid donde se dedica principalmente a los exámenes DELE en tareas de elaboración de pruebas, formación de examinadores y apoyo a centros de examen. Manuela es licenciada en Filología Hispánica (Universidad Autónoma de Madrid) y posee el máster en Enseñanza de ELE (Universidad Nebrija de Madrid).

Manuela Gil-Toresano is a Spanish teacher (Instituto Cervantes) and a teacher trainer. She is now working at the Language Certification Unit in the Instituto Cervantes' headquarters in Madrid where she is mainly involved in the development of the DELE exams and in the examiners' training courses. Manuela has a degree in Hispanic Philology (Universidad Autónoma de Madrid) and a master's degree in Teaching Spanish as a Foreign Language (Universidad Nebrija de Madrid).

Javier Fruns 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.

Javier Fruns is Head of the Unit of Language Certification of the Instituto Cervantes. He has participated in the design and preparation of the DELE and SIELE exams, as well as in the preparation of courses for examiners. He regularly teaches courses on language assessment at universities and official language schools. Javier has a Doctor of Letters from the University of Massachusetts and a degree in Hispanic Philology from the Universidad Autónoma de Madrid.

Latin America: An initial exploration of current language policies, national and local tests in Latin America

Nayibe Rosado Mendinueta

THU-27th

This presentation will share perspectives from different Latin American contexts: Brazil, Colombia, Chile, Mexico and Argentina. It will especially focus on the contextualization process that each country has undergone to design, and/ or to adapt global and governmental language policies to the design of localized language tests. The report is based on archival work and informal interviews conducted by LAALTA members in each of these contexts. The presentation aims to start a conversation with fellow academics about language testing in Latin America concerning language abilities in official, native/indigenous, and foreign languages, and also the policies for language testing that have been adopted in specific countries throughout the region. The conversation will also include government policies concerning other stakeholders (i.e citizenship tests and how they are operationalized) and whether or not these policies are linked to specific national or local tests. In relation to tests, it will report on official language tests delivered in the country (including where and how they are delivered, test taker target audience, intended test use, etc.).

Nayibe Rosado Mendinueta is the head of the Spanish Department at Universidad del Norte in Barranquilla, Colombia. She is currently the president of LAALTA. Her general research interests reside in the intersection of student learning and teacher learning and how language affects the construction of reality in contexts such as classrooms and in other institutions that surround us. Nayibe also studies curriculum and assessment as sites for exploring possibilities that support both student and teacher learning and development. Current research is strongly supported by action research: Academic literacy development and assessment, teacher identity, visible pedagogies to enable learning, and curricular development.

East Asia: Assessment Design for Language Teaching in Vietnam

Huu Nguyen

FRI-28th

Together with the comprehensive reform in education, since 2010, different activities have been implemented to enhance the teaching of foreign languages in Vietnam. In 2014, the Ministry of Education and Training of Vietnam issued Circular No.1/2014 stipulating the Framework of references for foreign language competences for Vietnam (the CEFR-VN), which was developed based on the CEFR. The CEFR-VN includes 6 levels of from Level 1 to Level 6, corresponding to Level A1 to Level C2 of the CEFR. This presentation will discuss how the CEFR-VN was developed and used in the education system of Vietnam, focusing on its application in English language assessment to form an assessment design for Vietnam. For general education, the Ministry of Education and Training of Vietnam promotes formative assessment, still highlights the importance of testing, specially proficiency testing across levels from A1 to B1. For tertiary education and life-long learning, standards were formed following the CEFR-VN levels of from Level 1 to Level 5 and assessment activities were designed accordingly. With supports of international assessment and testing experts, assessment activities in English language teaching in Vietnam have made Vietnam a considerable case study for other countries when applying the CEFR in their contexts and provide significant information for scholars as well.

Huu Nguyen was assigned to be the director of the National Foreign Languages Project, Ministry of Education and Training of Vietnam in January 2017 and has been in her office since then. She has engaged in a wide range of national projects, professional conferences, education activities and has made contribution to the development of Vietnam's national foreign languages education. Huu Nguyen delivered lectures and worked for years at Vietnam National University where she gained knowledge base for teaching and research profession. She served as the Deputy Director of Language Testing and Assessment Center of University of Languages and International Studies (Vietnam National University, Hanoi), where she was involved in the development of the Vietnam Standardized Test of English Proficiency (VSTEP), which is now a national test format of Vietnam. Huu Nguyen has been a visiting scholar at the Institute of Education - Massey University. Her research interest includes language test development and validation, test specifications, formative assessment, ESP, CLIL and EMI.

European commission Presentation

European Union policy on language learning and Key data on teaching languages at school in Europe [F]

Anna Solé Mena

FRI-28th

For many years, the European Union has been pursuing a language learning policy that encourages all young people to acquire foreign language competences from an early age, so that by the end of secondary education they are capable to master two languages, in addition to the language of schooling. This is an important element in achieving the European Education Area, as it will enable them to further study, work or do research in other EU countries. Another aim is to change the mind-set of policymakers and teachers, inspiring them to adopt comprehensive language education policies and innovative and inclusive language teaching methods. This includes the development of language awareness in schools. In order to be able to monitor progress toward our objectives two initiatives are especially relevant: Firstly, the fifth and brand-new edition of the report Key data on teaching languages at school in Europe includes 51 indicators covering many different topics in the field of (foreign) language teaching in schools in Europe. While foreign languages are at the centre of the investigation, other languages are also considered (regional or minority languages, classical languages, etc.). The presentation will unveil some of the report's most interesting findings on issues such as the number of foreign languages learnt by students at schools, the most popular foreign languages learnt by students, transnational mobility of foreign language teachers and language support for newly arrived migrant students. The report uses four different data sources: the Eurydice network, Eurostat, the 2018 Programme for International Students Assessment (PISA) and the 2018 Teaching and Learning International Survey (TALIS) from the OECD. Secondly, in order to complement this comprehensive report with an evaluation of student's abilities, the European Commission has been supporting OECD in the development of the 2025 PISA for foreign languages assessment (English) – which OECD will present.

Anna Solé Mena currently works as Senior expert on multilingualism at the Schools and Multilingualism Unit in the Directorate General for Education, Youth, Sport and Culture of the European Commission. She previously worked in different positions at the European Commission since 2001, notably in international affairs, industrial policy, sustainable development and climate change. Author of "Multilingual from the cradle. Growing up with different languages", 2010; Main character in the documentary on multilingualism "The Power of Babel", 2012.

Platinum Sponsor Session

Yunus Emre Institute: Language Teaching and Language Assessment Activities of Yunus Emre Institute

Dr Cumhur Ersin Adigüzel

THU-27th

In the speech, the Yunus Emre Institute will be briefly introduced, with a description of the Institute's activities. Following that, the Institute's educational, cultural, and artistic activities will be discussed. The number of the Institute's centres across the world will be mentioned. The method by which determined standards in Turkish teaching are closely followed, as well as the method by which various studies are carried out to achieve standardisation, will be discussed, along with examples of these studies. The speech will then be continued by drawing attention to the importance of measurement and evaluation practices in the language teaching process. The Institute's sensitivity and care in measuring and evaluating the Turkish proficiency of individuals learning Turkish as a foreign language, using validated and reliable measurement tools, will also be mentioned. It will be emphasised here that the TYS receiving the ALTE Order of Quality is significant not only for the Institute but also for our country.

During the speech, it will be mentioned that, as the world develops and progresses at a rapid pace, the activities of measuring and evaluating language skills, which are an important stage in the language teaching process, try to adapt to this development as well. In addition, the significance of adapting current approaches and tools used in the process of measuring and evaluating language skills will be emphasised.

Panel Sessions

ALTE Panel: Association of Language Testers in Europe (ALTE): promoting, maintaining, and safeguarding quality in language testing

Anne Gallagher, Nick Saville, Vincent Folny and Waldemar Martyniuk

WED-26th

Assessing language ability is not a straightforward endeavour because of its highly complex area of investigation: human cognition and behaviour. Over the last two decades the Association of Language Testers in Europe (ALTE) has been working on the development, promotion, and implementation of quality standards in language testing. The result of this huge collaborative effort of a large group of experts – representatives of ALTE member institutions – is the ALTE Quality Management System (QMS) based upon a set of 18 Minimum Standards established through a democratic process of discussion and consensus. ALTE experts have by now conducted over 100 audits designed within the system contributing to the improvement of language testing services offered by a variety of institutions in a number of languages. This achievement has been noted by all groups of stakeholders in the field, including authorities in several countries increasingly referring to the membership in ALTE and the ALTE Q-Mark as a guarantee for high quality service – a development that enhances the symbolic power of ALTE but at the same time raises the level of responsibility of its members. This panel will review the intentions behind the development, the principles it has been based upon and the implementation procedures of the ALTE QMS. Challenges related to the increasing value of the ALTE membership and the use of the ALTE Q-Mark will be addressed highlighting the fact that in the end both are to benefit and serve not just the associated language testing institutions but the test users, with the human beings under investigation in the first place.

Anne Gallagher worked as a lecturer of French, German and Irish for specific purposes and for over 20 years as Director of the Language Centre in Maynooth University, where she was appointed Director

of the Centre for Irish Language Research, Teaching and Testing and Head of the School of Celtic Studies until retirement in January 2021. Currently a member of the ALTE Board of Trustees, she has also served on a number of committees in Ireland and abroad focusing on the teaching and promotion of languages, and the digital transformation of the regions of Europe. In 2008 she was honoured by the French Government in recognition of her services to plurilingualism.

Dr **Nick Saville** is Director of Thought Leadership, English for Cambridge University Press & Assessment, and is Secretary-General of the Association of Language Testers in Europe (ALTE). His research and policy interests include: English language learning and plurilingualism; Learning Oriented Assessment (LOA) and the integration of learning with assessment (ILA); EdTech combined with EdAI for the automation of language assessment; language assessment literacy; and ethical frameworks in language education. He co-edits the *Studies in Language Testing* series (CUP).

Vincent Folny is in charge of innovation applied to French exams in France *éducation internationale* (FEI). He develops new exams, and he is involved in new developments of DELF-DALF and TCF. He is currently working on the development of an automated scoring of essays for the *Test de connaissance du Français* (TCF). Vincent FOLNY is actively involved in the definition and implementation of ALTE's quality management. He chairs the ALTE Standing committee, conducts regulatory audits in Europe and train auditors.

Waldemar Martyniuk is Professor at the Institute of Polish Language and Culture for Foreigners of the Jagiellonian University in Krakow, Poland. He holds a Ph. D. in Applied Linguistics and is a teacher trainer, author of several textbooks, curricula, and testing materials for Polish as a foreign language. Visiting Professor and lecturer at several universities in Germany, Switzerland (Basel), and in the USA (Stanford University). Seconded to the Council of Europe, Language Policy Division 2005-2006, and 2008-2013 as Executive Director of the European Centre for Modern Languages in Graz, Austria. Chair of the Board of Trustees at the Association of Language Testers in Europe (ALTE).

Implementing mediation: from theory to practice [F]

Joaquín Manuel Cruz Trapero, Adolfo Sánchez Cuadrado, Olga Naranjo Travieso and María Luz Castro Pérez

THU-27th

The publication of the Companion Volume (CV) has sparked off an intense debate on mediation in different contexts. From the theoretical standpoint, the conceptualization of mediation is still challenging. From the practical point of view, integrating mediation in curricula and language assessment poses different questions and challenges. We propose a debate at three different levels.

First, we will deepen into the notion of mediation, this complex and multi-faceted mode of communication that complements the communicative competence construct for foreign/second/additional language learning. To this end, how mediation intertwines with the other modes of communication, together with its specificity in terms of strategies and micro-skills, will be addressed.

Next, we will discuss how mediation can impact a particular context. Following the publication of the CV, Spanish legislation made it compulsory for Official Language Schools (OLS) to include mediation in their proficiency tests and in their curricula. We will discuss the two parallel systems used in the OLS of the Canary Islands, one targeting the teaching and progress assessment of mediation during the academic year, the other targeting the assessment of mediation in proficiency exams.

Finally, we will analyze the challenges that mediation poses in the development of scales for its assessment. We will discuss whether mediation should be assessed individually or embedded with other skills.

After this theoretical introduction, we will open a debate for participants to share their own questions and experiences from theoretical and applied perspectives. A reflection on the following questions will be sought: What is the perception of the implementation of mediation in language assessment and testing so far? To what extent are efforts across institutions, regions and/or countries being synchronized? What backwash effect is testing mediation having on teaching mediation? What tools are needed by test developers to include mediation into their tests?

Dr **Joaquín Cruz Trapero** is tenured at the Center for Higher Studies in Modern Languages of the University of Jaén, Spain, where he

has been in charge of language test design and validation since 2011. He has participated in multiple projects and conferences on test design and validation and he is currently the director of the international course on Language Test Design and Validation at its university.

Dr **Adolfo Sánchez Cuadrado** holds a BA in Translation and Interpreting, an MA in Foreign Language Pedagogy and a PhD in Linguistics. He is a lecturer in Linguistics at the University of Granada, and has also taught at the University of Delaware USA, at the Modern Language Centre of the University of Granada (where he was a member of the testing committee in charge of developing the *CertAcles* Spanish accreditation exam), and at University College London. His research interests lie within the fields of pedagogical translation, linguistic mediation, cross-linguistic cognitive grammar, and language teacher training.

Olga Naranjo Travieso has a degree in Translation and Interpreting from the University of Granada. She worked for the International Office of the University of Las Palmas de Gran Canaria, where she was in charge of International programs. In 2004 she joined the Official Language Schools (OLS) of the Canary Islands, in which she has developed several roles, including that of Director. Since 2018 she has been OLS coordinator in the Education Department of the Regional Government of the Canary Islands. In such capacity, she has been involved in the implementation of Mediation in the curriculum of the OLS of the Canary Islands.

M.^a Luz Castro Pérez has worked in *Escuelas Oficiales de Idiomas* in The Canary Islands since 1996 as a teacher, Head of Studies and Head of the English Department. Since 2016, she works for the Regional Government of The Canary Islands' Education Department, coordinating the area of foreign language learning. Among her responsibilities, she has coordinated the development and examining of the language proficiency tests, which are related to the CEFR's scales of proficiency (A2-C2), for the nine languages taught in this Autonomous Community. She has also coordinated the design of analytic scales to evaluate written and oral skills. Since 2020, she has been working in mediation exams and in its assessment scales.

Acreditación de la competencia comunicativa en lenguas minorizadas y contextos de diversidad lingüística

[DI]

Paula Elosua, Josep Ma. Montcada, Cristina Pérez-Guillot y Iñaki Villoslada

THU-27th

En comunidades históricamente bilingües sujetas a procesos de revitalización de sus lenguas minorizadas la certificación de la competencia comunicativa tiene características distintivas propias. El hecho de que el objeto de la certificación sea en muchos casos la lengua materna y/o la lengua de escolarización y, que el proceso de certificación sea gestionado por organismos públicos, definen un entorno diferente a aquél cubierto por la certificación en una segunda lengua. A las cuestiones psicométricas generales que se gestionan desde espacios públicos que han de adaptarse a los requerimientos metodológicos exigidos para alcanzar los estándares mínimos de calidad de sus procesos de acreditación, se unen la necesidad de transparencia y rendición de cuentas a las que estamos obligados, y entornos sociolingüísticos complejos y diversos en los que conviven dos lenguas que comparten estatus legal. En este panel, presentamos las experiencias de cuatro comunidades históricamente bilingües (Cataluña, Galicia, País Vasco y Valencia) en la gestión de los procesos de acreditación del catalán, gallego, euskera y valenciano. Cada una de las panelistas expone el contexto sociolingüístico en el que se llevan a cabo las pruebas de acreditación, su desarrollo histórico, los problemas y cuestiones de índole psicométrica, organizacional, económica, social e incluso política a los que se enfrentan en su quehacer diario, y el modo en que responden a ellos. El objetivo es crear un debate abierto en el que compartir experiencias, exponer líneas de trabajo y mostrar ante la comunidad internacional las significativas particularidades a los que nos enfrentemos en la acreditación de la competencia en nuestras lenguas nativas.

Additional Authors: Julia Zabala, ACLES y Olga López, Xunta de Galicia

Ph. D. Paula Elosua is professor of psychometrics in the University of the Basque Country. She has published over 100 papers on test assessment, test adaptation, validity, differential item functioning, score equating, IRT and SEM models. She has served as Secretary-General of the International Test Commission (2018-2022), and currently is president-elect of the IAAP assessment division.

Josep M.ª Montcada is currently in charge of language test design and language evaluation at the Department of Education of the Generalitat de Catalunya. He coordinates the development of high-stakes certification tests for the Official Language Schools (EOI) in Catalonia in 8 different languages and at levels B1, B2, C1 and C2. He trained in language assessment at Lancaster University and has attended numerous seminars and workshops on this topic, both as a participant and as a speaker.

Ph. D. Cristina Pérez-Guillot, Senior Lecturer at the Universitat Politècnica de València until september 2022. Director of the Applied Linguistic Department (1998 -2005). Director of the Language Center of the UPV from its creation in 2005 to May 2021. President of ACLES (Spanish Association of Language Centres in Higher Education) (2003- 2021). She has contributed to the development and publication of the ACLES Exam Accreditation Model, in January 2011, She is member of the Accreditation Committee as expert member.

Iñaki Villoslada, Bachelor of Basque Philology (University of Deusto) and degree in Law (UNED). Professor of Basque in the Euskaltegis of San Sebastian (1982-1990). He was evaluation technician in the IVAP for nine years, for the development of language profiling tests. He worked for the Basque Service of Health from 2000 to 2018, performing both standardization and translation tasks. He is currently head of didactics at the Institute for the teaching of Basque (HABE). In addition, he has written some articles on Basque rhetoric and has translated several history books on request.

PISA 2025 Foreign Language Assessment Framework: What matters for successful language learning and how to assess it [F]

Nick Saville, Angeliki Salamoura, Catalina Covacevich, Martin Robinson and Hanan Khalifa

FRI-28th

Proficiency in more than one language is a key asset for study, employability and cross-cultural communication in today's interconnected world. As a result, educational authorities worldwide are investing significant resources in foreign language teaching and learning (OECD 2020). Such investment, however, understandably raises important accountability questions: Are their students

achieving the expected language learning goals? Are their policies and instructional methods effective? Where should they concentrate their efforts with limited resources? The PISA 2025 Foreign Language Assessment (FLA) and Framework, developed by a partnership between OECD and Cambridge English, were designed to provide evidence that will answer these fundamental questions.

In language education, frameworks typically describe what is important to be mastered, taught and assessed at different proficiency levels. Together with standards-based assessment, frameworks are increasingly used by educational authorities as policy and accountability instruments (Hudson 2012). In this vein, the PISA FLA Framework defines foreign language proficiency and use; the skills and competencies necessary to use a language; and the cognitive, social and cultural factors which influence successful language learning. It also describes how these competencies, skills and factors will be assessed and reported to provide the required evidence that will guide education policy decisions and ultimately the improvement of language education. The PISA FLA Framework builds on the CEFR descriptive scheme (Council of Europe, 2020) and on Weir's (2005) socio-cognitive model of test validation. PISA 2025 will test English as a foreign language, and progressively cover other languages.

The panel will discuss the challenges inherent in the design and development of this large scale, international assessment, and address the following questions: Why does foreign language learning matter in 21st century education? What factors influence successful language learning? How to assess and report language learning outcomes? How will this information help guide education policy and improve future outcomes?

Dr Nick Saville is Director of Thought Leadership, English for Cambridge University Press & Assessment, and is Secretary-General of the Association of Language Testers in Europe (ALTE). His research and policy interests include: English language learning and plurilingualism; Learning Oriented Assessment (LOA) and the integration of learning with assessment (ILA); EdTech combined with EdAI for the automation of language assessment; language assessment literacy; and ethical frameworks in language education. He co-edits the Studies in Language Testing series (CUP).

Dr Angeliki Salamoura is Head of Operational Research at Cambridge English and leads research into the quality and validity of learning and assessment products. Angeliki also has extensive experience in Learning Oriented Assessment, integrated learning and assessment, education reform, impact and the CEFR. In this context, she has led numerous learning and assessment projects in the K-12 and vocational sectors in Europe, Asia and South America. She is one of the contributors to the PISA 2025 Foreign Language Assessment Framework.

Catalina Covacevich is an education policy analyst at the OECD, specialized in the design and implementation of large scale assessments at the national, regional and international level. She is currently working on the development of the PISA 2025 optional Foreign Language Assessment- English as well as supporting the design and implementation of PISA. She speaks English, Spanish, and some French.

Martin Robinson is an Assistant Director at Cambridge English leading the development of Cambridge automated, adaptive assessments, including Linguaskill and the Cambridge benchmarking service. Martin also leads the development of the content and scoring for the 2025 PISA FLA tests. Before joining Cambridge, he gained extensive experience in language teaching and school management in Spain and Japan. He holds an MA in Applied Linguistics, specialising in language assessment, from the University of Reading, and is qualified in PRINCE2 and MSP.

Dr Hanan Khalifa is a leading international expert in learning, teaching and assessment. Since 1993, she developed national examinations, validated international assessments, and led the alignment of local and international curricula and examinations to international standards, e.g., the CEFR. Her current work/interest is in the impact of reform initiatives, sustainability, and learning and assessing Arabic as L1. Dr Khalifa is an accomplished author with cited publications, an International Education awards winner, and a chair of education advisory boards.

Member Showcase Presentations

HABE: HABE C1: Evidencias y garantías de validez en la acreditación de la competencia comunicativa en euskera [DI]

Paula Elosua, Iñaki Villoslada y Enara Azkue

WED-26th

HABE C1 es una prueba para la acreditación de la competencia comunicativa en euskera en el nivel C gestionada desde el año 2017 por HABE (Helduen Alfabetatze eta Berreuskalduntzerako Erakundea). Su construcción, aplicación y corrección se ajustan a las recomendaciones de ALTE y del MCER.

En esta presentación nos centramos en el análisis de dos aspectos de la validez de las puntuaciones de HABE C1 que analizamos desde el punto de vista del concepto de validez propuesto por Messick y postulado por Kane. El objetivo es examinar las evidencias y garantías de validez relacionadas con las inferencias de explicación y generalización para las pruebas de comprensión escrita y oral. En el estudio sobre las inferencias de explicación analizamos el funcionamiento diferencial del ítem con relación a una variable criterio de alta relevancia en la evaluación de la competencia en Euskera: el origen de los candidatos. Teniendo en cuenta que las diferentes variantes del euskera tienen singularidades propias, el objetivo es comprobar que ninguno de los grupos lingüísticos se vea perjudicado por el contenido y/o redacción de los ítems.

Dentro de las inferencias de generalización nos centramos en los estudios llevados a cabo para definir e implementar un programa de equiparación de puntuaciones que tenga en cuenta los condicionantes externos a los que está sometida HABE C1. Los resultados avalan la utilización de un diseño de equiparación de grupos equivalentes. En un entorno históricamente bilingüe, la acreditación de la competencia en la primera lengua es un hecho a veces difícil de encajar. Las investigaciones y desarrollos que estamos llevando a cabo con HABE C1 persiguen alcanzar una prueba válida que se ajuste a los estándares de calidad internacionales más estrictos, y al mismo tiempo permitan mejorar su representación social.

Dr. Paula Elosua es catedrática de psicometría. Ha dirigido varios proyectos de investigación en concurrencia competitiva, y ha publicado más de 100 trabajos sobre evaluación por medio de pruebas, adaptación de test, validez, funcionamiento diferencial del ítem, equiparación de puntuaciones, modelos TRI y SEM. Ha sido secretaria general de la International Test Commission (2018-2022), y es presidenta electa de la división de evaluación de la IAAP.

Iñaki Villoslada es licenciado en Filología Vasca por la Universidad de Deusto y en Derecho por la UNED. Profesor de euskera de 1982 a 1990. Fue Técnico de evaluación del IVAP (1991-2000) y técnico de euskera de Osakidetza hasta 2018. Es técnico de euskaldunización en HABE donde ha desempeñado el puesto de responsable de Área de Evaluación y Acreditación y es actualmente responsable del Servicio de Euskaldunización. Ha escrito artículos de retórica vasca y ha traducido varios libros de Historia.

Enara Azkue es licenciada en Filología Vasca por la Universidad de Deusto. Traductora hasta 2015, año en el que empieza a desempeñar el puesto de técnica de normalización lingüística en la Diputación Foral de Gipuzkoa. Actualmente es técnica de euskaldunización en HABE, en el Área de Evaluación y Acreditación, en el nivel C1.

Cambridge University Press & Assessment: How can we make language assessment fit for the future? Current trends and future opportunities [D]

Angeliki Salamoura

THU-27th

It is fair to say that change has been the only constant in the education world in the past two decades. Language assessment has not been an exception, and in this paper I will discuss the multiple and, in some instances, antithetical trends that currently influence the field and shape its future.

On the technological front, AI advancements have allowed the development of assessment practices such as remote proctoring, adaptive testing and automarking. AI has also been used to create more personalised, non-linear learning approaches (OECD, 2018) which could allow individualised opportunities to practice or receive targeted feedback. On the theoretical front, we have witnessed a

growing consensus that assessment and learning are two sides of the same coin (e.g. learning oriented assessment, Turner and Purpura, 2016; Jones and Saville, 2016), as well as a broadening of the constructs entailed in language use to include, for instance, mediation, online language and plurilingualism (Council of Europe, 2020). The use of AI in applications such as automarking has also brought scepticism about its effectiveness and validity or has raised questions about ethical data use (cf. The Ethical Framework for AI in Education, 2021). Finally, there has been a gradual evolution of perceptions about the nature and role of English in the ELT world. In areas such as pronunciation a clear move away from native speaker norms towards intelligibility is taking place (Issacs, 2018) whereas the role of 'standard' English is being questioned in some contexts (e.g. Tajeddin et al., 2018).

Taking the perspective of an assessment provider, I will discuss the challenges inherent in these trends for ensuring fit-for-purpose and fit-for-use assessments, but also explore the opportunities opened up for making assessment more flexible, accessible and tailored to stakeholders' needs. I will illustrate with examples from Cambridge English assessment products.

Dr Angeliki Salamoura is Head of Operational Research at Cambridge English and leads research into the quality and validity of learning and assessment products. Angeliki also has extensive experience in Learning Oriented Assessment, integrated learning and assessment, education reform, impact and the CEFR. In this context, she has led numerous learning and assessment projects in the K-12 and vocational sectors in Europe, Asia and South America. She is one of the contributors to the PISA 2025 Foreign Language Assessment Framework.

telc gGmbH: Standardisation On-screen: Ensuring Exam Equity Across 24 European Languages [D]

R. Sean McDonald

THU-27th

For language assessment organisations, equivalence across languages is vital to maintain the validity of assessments when delivering on the global scale.

In this showcase presentation, discover how telc - language tests, based in Germany, are achieving their goal of creating valid,

reliable, and standardised assessments across 24 languages, all with adherence to the Common European Framework of Reference for Languages (CEFR). More importantly, learn how we can ensure equivalency across the 24 languages.

I will demonstrate how we are able to:

- Expand our exam offering to reach more regions and candidates
- Increase efficiency and reduce the logistical costs of delivering assessments
- Standardise assessments across a large and varied geographical area.

Through an innovative approach to on-screen testing that covers all four of the core language testing areas, including reading, writing, speaking and listening, we work to deliver reliable, reputable and effective assessments that demonstrate linguistic capability, replicating as close as possible real-world scenarios. This includes using the latest in item/question creation technology to ensure the candidate experience is enriched with detailed rich media, such as audio and video, to replicate those real-world situations.

I will share how we performed user research to ensure our tests are fit for purpose, and how this research fed back into test development. Finally, learn how technology allows teams of Subject-matter experts (SMEs) to work to achieve standardisation even in remote locations, and to cooperate securely across continents.

R. Sean McDonald manages on-screen test development at telc gGmbH in Bad Hamburg, Germany. He works closely with technical departments in the development of on-screen testing to ensure that telc tests are valid and fully aligned with telc's paper examinations. Sean has University degrees from the University of California and University of Washington in Seattle. He has been with telc gGmbH for over 10 years and is a regular speaker at assessment events with a particular focus on innovation and emerging technologies.

CNaVT: Reporting test scores in a more insightful way [DI]

Manon Willems and Goedele Vandommele

THU-27th

Reporting test results in an insightful and clear manner is an important aspect of test validity (ALTE Manual for Language Test

Development and Examining, p.47). For this reason, the Certificate Dutch as a Foreign Language (CNaVT) is always looking for ways to improve communication around test results. In this presentation, we will present our improvements and their concretization, following a seven-step-model for the development of score reports (Hambleton & Zenisky, 2013).

We carried out the first three steps to determine what our score report should look like:

1. A needs analysis in the form of an inventory of feedback received over the last five years.
2. The identification of intended audience(s), their background and needs (e.g., mainly candidates and their teachers, and secondly, parents, school management and admission officers).
3. A review of available literature and report samples with insightful design in search of elements to include in our score report.

Based on the information gathered, we decided on four improvements:

1. Using a more intuitive scale (as compared to the candidate's Rasch measure), in order to prevent misinterpretation.
2. Reporting subscores to provide insight into candidates' strengths and weaknesses and help them progress.
3. Visualizing our scores by combining different presentation modes (Ryan, 2006).
4. Including supporting material, such as an annotated score report (Goodman & Hambleton, 2004) that can be tailored to the needs of different stakeholders.

In the following months, we will carry out Hambleton and Zenisky's (2013) remaining four steps, which include draft reports, field tests, a redesign phase and the continuous monitoring of the final draft. In this showcase, we will present our findings and our reflections that led to decisions. Feedback from the audience will serve to finalize our improved communication.

Additional Authors: Jonas Wynants, CNaVT and Anne van Asseldonk, CNaVT.

Manon Willems obtained her bachelor's degree in Linguistics from the Radboud University in Nijmegen, the Netherlands. During her master's degree on the Utrecht University, she specialized in multilingualism and language acquisition. Since February 2022, Manon is test developer for the Certificate Dutch as a Foreign Language (CNaVT). Her focus is the PROF-exam, which is an exam for people who need Dutch for their job, especially in the healthcare sector or the administrative sector.

Goedele Vandommele is postdoctoral research fellow and current director of the Certificate Dutch as a Foreign Language (CNaVT) at the Centre for Language and Education (KU Leuven). The CNaVT certifies language proficiency of Dutch as a foreign language worldwide, using task based and domain specific exams related to the Common European Framework of Reference for Languages (CEFR). She is also part of two consortia with a focus on (language) assessment: Steunpunt Centrale Toetsen in Onderwijs and IELS (International Early Learning and Child Well-being Study). Her main research topics are (second) language acquisition, effective instruction and language assessment, with a specific focus on the adolescent and vulnerable language learners, such as young newcomers in Flanders.

Charles University Prague: Realignment of the Czech Language Certificate Exam to the CEFR: Processes, issues and implications [F]

Martina Hulešová

THU-27th

Since the generally accepted Messick's view of validity is related to the adequacy and appropriateness of interpretations of test results and decisions and social consequences based on the scores (Messick, 1990), the emphasis on transparent interpretation of the scores and their meaning is extremely important. It also concerns the aspect of comparability of scores from different test forms and testing sessions. Therefore, aligning exams to frameworks, being this local frameworks, or internationally recognised frameworks such as the CEFR, seems to be one of the crucial and necessary steps on the way towards fairness and justice in language testing and assessment.

In this paper, I outline the process of alignment of the Czech Language Certificate Exam (CCE) to the CEFR, starting in 2014 and finishing in 2023. It turned out to be a complex process that affected a suite of exams at five CEFR levels, both productive and receptive skills, and caused substantial changes in the test development, administration, and rating processes. Qualitative phases, such as the content analysis and the following content, specifications and rating criteria revision will be presented. Quantitative methods used for setting the standard and for the alignment to the CEFR-related interpretation of the results will also be shown. Limitations and challenges, related mostly, but not exclusively, to the fact Czech is one of the less widely spoken languages, will be outlined, too.

Messick, S. (1990). Validity of Test Interpretation and Use. Educational Testing Service, Princeton, N.J. Accessed Sep, 22, 2022 from <https://files.eric.ed.gov/fulltext/ED395031.pdf>

Martina Hulešová graduated in Spanish Philology at Charles University (Prague, Czech Republic) and later obtained her MA in Language Testing at Lancaster University (UK) and PhD in Didactics of Foreign Languages at Masaryk University (Brno, Czech Republic). Her research interests include standard setting, validation and training in language testing and assessment. She works as quality manager for the Research and Test Centre at the Institute for Language and Preparatory Studies (Charles University) and as a freelance she provides consultancy for other projects related to language assessment, Czech sign language, diagnostic tests and others. She is a member of the Executive Committee at the EALTA (European Association for Language Testing and Assessment).

Foreign Language Centre, University of Pecs: Does the mode of delivery influence test takers' performance? A comparative analysis [D]

Róbert Márcz and Réka Werner

THU-27th

The ECL Language Examination Centre offers standardized, general-purpose and monolingual foreign language proficiency exams consisting of four parts: reading comprehension, listening comprehension, written communication, and oral communication. The exam was originally offered in a paper-based format. In 2020 the ECL International Centre developed a computer-based version of its language exam with two modes of delivery: candidates can take computer-based language exams at designated examination sites and at home.

In order to find out whether the medium (computer) and the circumstances (proctored online exam) have any influence on test takers' performance, a comparability study was conducted. German and English language tests measuring three skills (reading comprehension, listening comprehension and written communication) that had been used previously during live exams were converted into the online format for candidates who volunteered to take a mock online exam. The composition of the mock exam participants broadly mapped the profile of the ECL

test population in terms of age, gender, place of residence, foreign language ability and language level. In the case of the B2 level exams, more than 500 candidates completed both paper-based and online tests in English, with close to 200 candidates also completing both tests in German.

The statistical analyses of the previous live exam results and those of the online mock exam were compared. In order to test the null hypothesis that there was no significance difference between the performances of candidates taking the test in the two modes of delivery, the Independent Samples T Test was applied. The results of the comparative analyses are presented.

Róbert Márcz has a MA degree in English and History at the University of Pécs. Head of the English Department of the ECL International Language Examination Centre, responsible for item writing, test development and examiners' training. Obtained a PhD in the Doctoral Programme in English Applied Linguistics and TEFL/ TESL at the University of Pécs (2019). Title of dissertation: The Washback of Language Exams on Classroom Practice in Hungary in a Complex Dynamic System Framework. Member of the Hungarian Accreditation Board for Foreign Language Examinations.

Réka Werner graduated having an MA degree in German Language at the University of Pécs. Head of the ECL International Language Examination Centre. Worked for the Educational Authority as the coordinator of the German team elaborating the language tests for the national foreign language competency assessment.

Le Français des Affaires: Combine strengths to provide high quality digitized language tests and examinations [D]

Dominique Casanova and Bert Wylín

FRI-28th

Le français des affaires has been an early user of digitization in language testing. The first digitized version of the Test d'évaluation de français – TEF – was a homemade system working mainly offline. But it was tailor-made for the test and not able to satisfy new needs. That is the reason why Le français des affaires decided to work with e-assessment platform vendors when it renewed its IT system.

Such a collaboration has lots of advantages if and when the provider listens to the customer's needs. First of all, the language tester remains focused on his core business, not having to deal directly with efficient implementation or technical problems. Second, a professional e-assessment platform comes with a large list of possibilities which would take time and money to develop by its own. And it continues evolving and updating, following and even anticipating new findings in the field of evaluation, like psychometrics, proctoring, AI-generated items or adaptive tests.

Conversely the e-assessment platform vendors also benefit from the collaboration with demanding customers. It helps them to validate the relevance of new ideas, the expected level of requirements and to identify needs which could be common to different customers. Custom made is not always the best path, it helps to check how generic certain features are; it often makes them better.

In this presentation, Le français des affaires and Televic education will provide feedback on the way they collaborate to enhance both the quality of the digitized language tests and the e-assessment platform and illustrate it with concrete examples.

Dominique Casanova is Head of scientific development at Le français des affaires. He started working in the field of language testing in 2005, when coordinating the development of the electronic version of the Test d'évaluation de Français. In 2013-15 he managed the renovation of the IT system of Le français des affaires. He currently supervises psychometrics and projects involving machine learning and NLP. He is one of the chairs of the ALTE Special Interest Group on Technology in Language Assessment.

Bert Wylín has both an academic and a business profile. He worked at the KU Leuven since 1993, leading the Education Innovation Centre, for integrating EdTech in university education. In 2001, he founded a KU Leuven university spin-off, now Televic Education, developing and servicing educational technology projects and products. Today, Bert's focus is on the content side of e-assessment: how does technology best serve the didactic and pedagogical needs of content and skills measurement and assessment?

TestDaF-Institut: New test format – new research agenda: An overview of the technology-related research at g.a.s.t. [D]

Prof. Dr. Anastasia Drackert

FRI-28th

The portfolio of the Society for Academic Study Preparation and Test Development (Gesellschaft für akademische Studienvorbereitung und Testentwicklung, g.a.s.t.) includes language tests and scholastic aptitude tests, as well as online German courses for international study applicants. In the last years, digital versions of our tests have been introduced, e.g. in late 2020 the digital Test of German as a Foreign Language (Test Deutsch als Fremdsprache, TestDaF). The introduction of the new tests has also called for new validation studies.

The presentation will give an overview of the research being conducted at our institution on the digital TestDaF. In particular, it will focus on the different aspects of validation and quality assurance which relate to all three strands of the conference. In regard to strand 1 (Fit for the Digital Age), we will report on the research aiming to incorporate automarking of test-takers' short answers to the listening comprehension tasks and discuss the challenges still to overcome to make a fully automatic assessment possible. In regard to strand 2 (Diversity and Inclusion in Language Assessment), we will present challenges and promises of accommodations that we offer to the test-takers with special needs in digital test environments. Finally, in regard to strand 3 (Implementation of Frameworks) we will report on the online standard-setting conducted to align the revised digital TestDaF to the CEFR. Overall, the goal of the presentation is to show how we use technology to improve our assessments and make them fair and accessible for German learners across the globe.

Additional Authors: Sonja Zimmermann, Leska Schwarz and Anja Peters.

Anastasia Drackert is Director of Research at the Society for Academic Study Preparation and Test Development. She is also a professor in language testing and digital learning at Ruhr-University Bochum. She received her PhD at Georgetown University in 2015 with the dissertation "Validating Language Proficiency Assessments in Second Language Acquisition Research. Applying an Argument-Based Approach".

CVCL Universita per Stranieri Perugia: The CELI corpus: a new tool to analyse written L2 Italian and to inform language testing and assessment [D]

Fabio Zanda and Danilo Rini

FRI-28th

While corpora have been a reference resource in the fields of language acquisition and language assessment for almost thirty years, the dearth of learner corpora containing reliable learner proficiency level data represents a major weak point when developing a new corpus-based test or revising an existing test (Barker et al. 2015).

In this presentation, we illustrate the main features of a new learner corpus designed to analyse Italian L2, the CELI corpus (Spina et al. 2022). The CELI corpus has been designed to systematically compile the written texts produced by candidates of Italian L2 who have passed the CELI exams at B1, B2, C1 and C2 proficiency levels of the CEFR. Over 3000 texts were included in the corpus, with a balanced distribution of tokens in terms of proficiency level, totaling ca. 600,000 tokens. The transcribed texts were fully anonymized, then lemmatized, pos-tagged, and manually processed. Furthermore, all texts are associated with metadata which include learner variables and task variables (Granger 2008). In order to allow online queries of words, word sequences and linguistic annotations, a web-based platform was made available on a CQPweb system (Hardie 2012). In terms of practical benefits, the use of learner corpora representing production from candidates in an examination setting enable test developers "to increase transparency, consistency and comparability in the assessment of L2 proficiency, and in particular, to inform, validate, and advance the way L2 proficiency is assessed in the CEFR" (Callies & Götz 2015:3). Hence, we present some of the potential applications which could be made of the CELI corpus, from developing corpus-informed tests to improving raters' training (cf. Gablasova 2021). The CELI corpus is a new tool containing reliable proficiency level data which aims to expand the array of resources available for future research of L2 Italian language testing and assessment studies.

References: Barker, F., Salamoura, A., & Saville, N. (2015). Learner corpora and language testing. In S. Granger, G. Gilquin, & F. Meunier (Eds.), The Cambridge handbook of learner corpus research (pp. 511–533). Cambridge: Cambridge University Press. Callies, M.,

& Götz, S. (Eds.) (2015). Learner corpora in language testing and assessment. Amsterdam: John Benjamins. Gablasova, D. (2021). Corpora for second language assessments. In N. Tracy-Ventura, M. Paquot (Eds.), *The Routledge Handbook of Second Language Acquisition and Language Testing* (pp. 45-53). London-New York: Routledge. Granger S. (2008). Learner corpora. In A. Lüdeling, M. Kytö (Eds.), *Corpus Linguistics. An International Handbook*. Volume 1 (pp. 259-275). Berlin: Mouton de Gruyter. Hardie, A (2012). CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17 (3), 380-409. Spina, S., Fioravanti, I., Forti, L., Santucci, V., Scerra, A., Zanda, F. (2022). Il corpus CELI: una nuova risorsa per studiare l'acquisizione dell'italiano L2. *Italiano LinguaDue*, 14(1), 116-138.

Additional Authors: Stefania Spina, Luciana Forti and Irene Fioravanti, University for Foreigners of Perugia.

Fabio Zanda is a PhD candidate in Linguistics and Language Teaching at the University for Foreigners of Perugia. He is currently developing a collocational knowledge test for learners of L2 Italian. He is particularly interested in Corpus Linguistics, Learner Corpus Research and issues related to the CELI (Italian Language Certification) exams, as well as language testing and vocabulary assessment studies.

Danilo Rini taught Italian as a foreign language both in Italy and abroad, before joining the CVCL (Centro per la Valutazione e le Certificazioni Linguistiche) of the University for Foreigners of Perugia, in 2005. There, he worked as a rater, examiner, and on the planning and development of the internal item banking system. He currently works as a test developer and reviewer for CELI (Certificato di Lingua Italiana) exams, as an examiner and trainer for examiners, and on the quality assurance of CELI exams through statistical analyses.

Sign Language Presentations

Assessing receptive skills development in deaf children who use Swiss German Sign Language as their primary language [DI]

Tobias Haug

THU-27th

Many sign languages are under-researched and under-resourced, this is even more pronounced in small countries where more than one sign language is used, like in Switzerland. In the German speaking part of Switzerland, there is a huge demand for instruments to assess the development of Swiss German Sign Language (DSGS) in deaf children in the (pre)school context. So far, no operational test for DSGS is available. In response to this need for DSGS tests, we developed and piloted two receptive skills tests targeting deaf children between 4-14 years (N = 37, M age = 9.00). One test is designed to assess the comprehension of morphological constructions (N items = 46) and the other test targets narrative comprehension (N items = 17). Both tests use three- and four-option multiple choice items, with one correct option per item. Preliminary statistical analysis revealed good results to inform the revision of the tests. Cronbach's Alpha across all items was .94 and few items (n = 8) should either be removed or revised in both tests, using a CICT value of >.25 as a criterion for removal from the item pool. Additionally, there was a strong and statistically significant correlation between the raw scores and the chronological age of the children. These results are promising since it would be expected in developmental tests that older children outperform younger children. We also correlated the scores of both tests to see if they represent a similar construct. The results reveal also a strong and statistically significant correlation, which suggests that both tests tap into similar language skills. The tests are currently being revised and used in a new project that will collect additional data over a period of five years to norm the tests with deaf children using DSGS as their primary language.

Additional Authors: Franz Holzknacht, University of Teacher Education in Special Needs, Regula Perrollaz, University of Teacher Education in Special Needs.

Tobias Haug earned in 2009 his PhD from Hamburg University and 2017 his MA in Language Testing from Lancaster University. Since 2004 he has been the director of the sign language interpreter program at the University of Teacher Education in Special Needs (HfH) Zurich. Since 2018 he is also Professor for Sign Languages and Participation of the Deaf. He has been involved in a number of research projects related to sign language testing and assessment.

Common European Framework of Reference and Czech Sign Language, project 2019-2022 [DI]

Vladimir Simon and Denisa Lachmanova

THU-27th

The project Common European Framework of Reference linked to Czech Sign Language started in 2019 in the Czech Republic. After 4 years of research we are pleased to present three outcomes - a Framework of Reference for Sign Languages (FRSL), Reference Level Descriptors for Czech Sign Language A1-B2 (RLDCSL) and the website containing translation into Czech Sign Language. In the FRSL we created new communicative situations related to communicative language activities (reception, production, interaction, mediation) that fit more precisely to real-world communication in sign languages. We also dealt with chapters including communicative language competences (linguistics, pragmalinguistics, sociolinguistics). In the second document we described Czech Sign Language and edited chapters e.g. Overview of grammar and vocabulary of Czech Sign Language, Socio-Cultural knowledge and abilities, Themes of communication. What were the most difficult decisions during the project? How did we design new scales? How did we deal with validation? We want to share our experience and inspire other teams to dedicate their time to this topic.

Additional Authors: Radka Novakova, Charles University, National Pedagogical Institute of the Czech Republic, Romana Petranova, National Pedagogical Institute of the Czech Republic, Lucie Stadlerova, National Pedagogical Institute of the Czech Republic.

Vladimir Simon is Deafblind, native signer of Czech Sign Language, supervisor, a language proofreader and a translator from written Czech to Czech Sign Language. He studied at the Institute of Deaf Studies, the Faculty of Arts of Charles University in Prague.

His interests include the linguistics of sign language, deaf culture, cognitive linguistics, etc. He has been working in Pevnost – the Czech Centre of Sign Language since 2012 providing courses of Czech Sign Language. He has experience working with interpreters in linguistic-translational field.

Denisa Lachmanova is Interpreter Czech – Czech Sign Language. She has practiced as a interpreter since 2014. After completing her MA at the Institute of Deaf Studies at Faculty of Arts at Charles University in Prague in 2016, she started working there and is still a member of the academic team. She is also working in a team with both deaf and hearing colleagues in a Czech public television. She is a member of professional organization for interpreters (the Czech Chamber of Sign Language Interpreters).

Workshops

Considerations and challenges in setting up a retrospective DE&I review of a high-stakes item bank [DI]

Kasia Wojciechowska and Lottie Harrison

WED-26th

In recent years the focus on diversity, equity and inclusion (DE&I) in content development has gained necessary traction, and within assessment there is an increased focus on producing content that not only excludes sensitivities but also actively includes and represents all test takers. But what about the content that was produced prior to this shift and is still actively in use?

Pearson has made a public commitment to fostering diversity, equity and inclusion in all its content, and assessment is no exception. We want to ensure that we are directly putting these words into action and have conducted an extensive, retrospective review of an existing, high stakes item bank, in a way that seems to be unrivalled in language testing.

Using this review as a case study, we will showcase one way to turn DE&I guidelines into review-focused material and look at how we applied this to historical content in an active item bank. This review

was based on the publication of Pearson's new, in-depth Global Content Editorial Policy which formed the backbone of the content evaluation. We will look at the challenges and constraints posed not only by dealing with high stakes content but also the impact of a global product on DE&I. We will also explore building care into the project to ensure that the reviewers were supported in what can be emotionally challenging work.

We hope that in sharing this ground-breaking review, we can pave the way for many more like it, not only at large-scale publishers, but across all language assessments, to ensure that DE&I practices are acted upon, not only talked about.

Additional Author: Dr. Amina Douidi

Kasia Wojciechowska is a Senior Manager High Stakes Tests at Pearson Education, responsible for ensuring that all test development processes are executed to the highest standard and that test material is fit for purpose. She also contributes to the development of digital assessment products, which deliver results to customers using AI technology. She has BA in Teaching English as a Foreign Language from A. Mickiewicz University, Poland and MA in Applied Linguistics, ESP from Northumbria University, UK.

Lottie Harrison is an English Language Assessment Manager at Pearson Education. After many years in English language teaching, management and teacher education, she is now responsible for test development and supporting a range of assessment related activities. She has an MA in English Studies from the University of Hong Kong. She also provides language classes for refugees to the UK and has been involved in mentoring programs such as EVE, promoting equal voices in ELT.

Bigger Bang for Your Assessment Buck: How Language Counselling Enhances the Relevance of Language Assessment for All Language Learners [DI]

Erin McDonald and Olga Sarycheva

FRI-28th

In Nova Scotia, the path towards language learning begins with a language assessment called the Canadian Language Benchmarks Placement Test (CLBPT). This type of language assessment,

funded by both the federal and provincial governments, enables any Canadian immigrant aged 16+ to take part in a government-funded language program(s) to begin or continue to improve their English communication skills. Since 2011, Learn English Nova Scotia Testing and Counselling Centre (LENS) has built their expertise in delivering language counselling sessions following each assessment.

Language counselling sessions are tailored to the unique learning needs and goals of the learner. Interpretation is provided, if needed, to ensure that learners can fully understand the results and ask questions. The demands on the service have been shaped over time by a range of factors including the Syrian initiative, subsequent emergency arrivals and most fundamentally, the COVID pandemic.

Language education counselling has come to evolve to meet the emerging needs of our learners no matter where they are in the province through remote service delivery. We have also adapted our services to better meet the needs of youth and young adult learners who need to navigate not only language acquisition, but also how language skills impact the path to high school completion. Despite all these changes and challenges, Language Counselling continues to be based on the principles of Adult Learning, which emphasize the importance of self-knowledge, goal setting, and learner autonomy. We believe that having ready access to the CLBPT and up to date information about all language learning options gives our clients no matter their age, literacy, language level or short and longer term goals, some sense of control as they make plans to return to school, work and build new careers in their new country.

Erin McDonald has worked in the field of adult immigrant education since 1998. She has worked with the Canadian Language Benchmarks since 2004 and has owned and operated Learn English NS Testing and Counselling Centre since 2011.

Olga Sarycheva, LENS Language Assessment and Counselling Coordinator, has 27 years of experience working in the ESL/EAL field. She holds a Master's Degree in Applied Linguistics with teaching qualifications. She is also a certified Canadian Language Benchmarks Placement Test (CLBPT) and Canadian English Language Benchmark Assessment for Nurses (CELBAN) Speaking assessor. In 2021 she acted as the Regional Trainer for the Atlantic Provinces during the roll-out of the new assessment tool for newcomers in Canada – Canadian Language Benchmarks Placement Assessment (CLBPA).

Assessing lesser used languages: Papiamentu and Dutch as a Foreign Language [D]

Judith Richters and Janneke Helsloot

FRI-28th

The Dutch Ministry of Education, Culture and Sciences issued a tender for the development of language tests for Bonaire, Statia (St. Eustatius) and Saba (shortly the BES-islands, officially municipalities of the Netherlands). Bureau ICE gained this tender and is now developing language tests Papiamentu, Dutch as a foreign Language and English. The tests are taken by pupils from group 3 primary education to group 3 secondary education. The project also includes the development of a Student Monitoring System for schools (in Dutch Leerlingvolgsysteem (LVS)). The languages Dutch as a Foreign Language and Papiamentu (the local Creole language on Bonaire) are new to the islands when it comes to testing. Dutch as a foreign language is based on the CEFR, for Papiamentu a whole new framework of reference has been developed. Given the complex language situation on the three islands (Dutch and Papiamentu the official languages on Bonaire, English the official language on Saba and Statia) developing one SMS for all islands is a challenge. Not only when it comes to choosing the levels and frameworks, but also from a psychometric point of view; for some tests no more than 20 students are available at that level to participate in the pretest. In this workshop we give you more background on the decision to test a 'tiny' language like Papiamentu and Dutch as a foreign language instead of Dutch as a mother tongue, and we will explore all the possible pitfalls we encounter.

Judith Richters is product manager government projects at Bureau ICE. She studied Spanish, Dutch as a second language and Arabic. She has worked as a consultant in education on themes like language teaching and testing, language policy, intercultural education, social internship. In her current position she is responsible for the development of different tests for Dutch as a second language (immigrant exams), learning ability tests and a student monitoring system for the BES-islands.

Janneke Helsloot is manager product management at Bureau ICE. She studied Language and Culture studies. She has worked as a consultant on themes like educational methods, transfer of paper-based examination processes to digital examination processes, development of integration examens. In her current position she is

responsible for the development of paper-based and digital tests for primary education: IEP Endtest (reading, spelling and maths) and IEP student monitoring system (language, math), and self-assessments for social-emotional development, learning approach and creative skills

Scholarship Winners' Presentations

FRIDAY 28th APRIL

Overall Scholarship Winner's Presentation: Spoken Dialog System-Delivered Paired Discussion Tasks for Interactional Competence Assessment [D]

Nazlinur Gokturk

Research shows that paired/group discussion tasks may give test-takers unique opportunities to demonstrate their interactional competence (IC) (e.g., May et al., 2020). However, the use of these tasks in large-scale assessments remains limited due to the impracticality of task administration and the potential impact of interlocutor characteristics on performance (e.g., Artunc & Hart, 2020). Several studies suggest that this issue may be addressed by building Spoken Dialog Systems (SDSs) to deliver versions of paired/group discussion tasks where computers serve as interlocutors (e.g., Chukharev-Hudilainen & Ockey, 2021). However, due to the lack of validation research on such SDS-delivered tasks, much remains to be learned about their appropriateness for IC assessment.

This research aims to explore the use of IC features by high- and low-ability test-takers in performance on an SDS-delivered paired discussion task, along with the perceived appropriateness of the SDS as an interlocutor. Data collected from 30 test-takers (ESL students at a university in the U.S.) and three trained raters included

test-taker task responses, task scores, and semi-structured interviews. The task responses were qualitatively analyzed for IC features using a coding scheme adapted from Galaczi (2014), and the use of the features within the task and between ability levels was summarized through descriptive statistics. The interviews were also qualitatively analyzed following a two-cycle coding method (Saldaña, 2013).

Findings showed that 13 out of 18 IC features were observable in the task responses, and the use of most observed features differed systematically between the two ability levels. Test-takers generally perceived the system as an appropriate interlocutor for oral communication assessment, although they noted several system limitations that could affect their task performance. Overall, this study helps us better understand the construct of IC as assessed by SDS-delivered paired discussion tasks and offers implications for the design and validation of these tasks.

Additional Author: *Dr. Evgeny Chukharev-Hudilainen, Iowa State University*

Dr. Nazlinur Gokturk works for the Board of Education at the Republic of Türkiye Ministry of National Education. She obtained her Ph.D. in Applied Linguistics and Technology from Iowa State University. Her research interests focus on developing and validating technology-mediated language tests, especially those that aim to assess second language oral communication ability. At the Ministry, she is currently working as an academic coordinator for a national project on the alignment of language education in Türkiye with the CEFR CV.

Man and Machine: Match made in heaven? (Status: "It's complicated...") – Exploring Dutch Automated Essay Scoring for young learners' written output [D]

Geert De Meyer

Recent educational policies have set out that Flemish education will implement large-scale government-lead standardized testing, including Dutch reading comprehension and writing. This study focuses on the possibilities of Automated Essay Scoring (AES) to consistently, fairly and practically assess a large number of writing

products. AES systems automatically score a text using machine learning and by extracting linguistic characteristics from a text (Allen et al., 2016).

While research has shown AES can be used to assess writing, previous research has focused almost exclusively on one particular language (English) and genre (essays) mostly written for higher education purposes (Strobl et al., 2019). This exploratory study investigates the possibilities of AES to obtain reliable scoring for Dutch-speaking learners in the first stage of secondary education.

A corpus of 5,110 writing products of 2,613 pupils aged 13-14, based on six prompts, was holistically scored by 852 in-service and pre-service teachers, using pairwise comparison. This assessed corpus was used to train machine learning models and, as such, create a first AES system to assess Dutch writing products of learners in the first stage of secondary education. We experimented with two flavours of machine learning: a traditional feature-based approach and a deep learning one. For the first approach, all writing was processed with T-Scan (Pander Maat et al., 2014) to derive linguistic text characteristics including lexical and syntactic measures. The second deep learning approach relies on Dutch state-of-the-art pre-trained language models (Delobelle et al., 2020), which have been fine-tuned on the task of AES.

The predictions of the AES system will be discussed in relation to previous studies that have investigated the role of AES in assessing writing in education, as well as implications for the use of AES in the context of language testing, and zoom in on challenges? when working with data of young learners.

Additional Authors: Fauve De Backer and Orphée De Clercq (UGent).

Geert De Meyer is a PhD researcher at the Centre for Diversity & Learning (Department of Linguistics, Faculty of Arts & Philosophy - Ghent University) under the supervision of Prof. Dr. Maribel Montero Perez, Prof. Dr. Orphée De Clercq (LT3) en Dr. Fauve De Backer. He is researching the opportunities of automated writing evaluation for writing in Dutch for primary and secondary school pupils. His research interests are language assessment and second language teaching methodology. Prior to his PhD research he was a language teacher (Dutch and English) in a secondary school and a teacher trainer at Artevelde University of Applied Sciences (Ghent), teaching courses in second language teaching methodology and coaching student teachers.

Exploring anti-plagiarism tool effects in the assessment of academic reading-into-writing [D]

Valeriia Koval

One of the core competencies in integrated reading-into-writing is the ability to transform language from a source by employing effective paraphrasing. To assess this competence validly, raters need to know the input material well. Long and complex input may threaten the objectivity of the rating as it creates cognitive load for the raters who are expected to clearly differentiate between language taken from source text and language produced by test taker. Therefore, the assessment of this competence by raters is rather challenging (Gebriil & Plakans, 2014). A possible improvement of the assessment accuracy and efficiency can be attained by employing a tool that highlights copied passages from the source in the test taker's writing. In my study, I investigate how the application and efficiency of one such tool is perceived by raters. The research includes the analysis of Focus Group Interviews (FGIs) with five raters who shared their experience of rating with and without the tool. The data from these FGIs is analyzed thematically (Braun & Clarke, 2006). Apart from revealing the facilitative effects of the tool, the results provide insights into the effects of the raters' source text knowledge as well as their severity and strategies when rating with and without the tool. The in-depth investigation of the application of an anti-plagiarism tool during the rating of integrated writing provides an important missing element in the research on rater cognition. Moreover, this research provides evidences that application of an assisting tool contributes to a fair, objective and efficient assessment of academic reading-into-writing for placement or diagnostic purposes.

Valeriia Koval is a PhD candidate and a research associate at the University of Bremen since 2020. She holds a Master's Degree in applied linguistics from the University of Bonn, Germany. Her research focus is assessment of academic integrated writing.

Bias is everywhere? An investigation into DIF at the item, rater and task level [DI]

Christine Troussart Van Bulck

The Certificate Dutch as a Foreign Language (CNaVT) develops five exams each year to assess Dutch language proficiency. One of these, the B2 STRT exam, is taken by candidates around the world to gain access to Dutch higher education in Belgium and the Netherlands. This exam context leads to conflicting demands: the exam must consist of tasks that are authentic and relevant within the context of Western European higher education, while cultural (and other) bias must be minimized to ensure fair testing for all candidates. Bias research is thus an important aspect of the standard STRT exam development and management. CNaVT conducts annual DIF analysis in Tiaplus to investigate which items cause bias regarding age, sex, linguistic, cultural and educational background. These analyses show DIF for approximately 4% of items. Cultural and linguistic bias is found on language accuracy scores, whereas content items are overwhelmingly free from DIF. We suspect that the STRT exam items are not inherently biased, but that bias occurs because of rater inconsistency. (Troussart Van Bulck et al, 2022)

Because this hypothesis could not be tested in Tiaplus, the next step of our bias investigation was to carry out a DIF analysis in Facets to analyze the role of the rater. In addition, we investigated whether bias is also found at task level and if so, which contexts and/or linguistics tasks give rise to bias. For this analysis, we combined all datasets since 2017 into one large dataset (N = 3746; total tasks = 18) to detect previously unnoticed trends. In our presentation, we discuss the results of the Facets DIF analysis. We give an overview of DIF at task and rater level and their potential causes. We conclude by formulating a number of potential measures to avoid bias at task or rater level.

Additional Author: Goedele Vandommele (KU Leuven).

Christine Troussart Van Bulck obtained her MA in Linguistics at KU Leuven (Belgium). She currently works for the Certificate Dutch as a Foreign Language (CNaVT), where she develops task-based exams for teenagers. Her research interests include language assessment and diversity, linguistic variation... She recently authored a poster presentation on item-level bias in the CNaVT STRT exam.

Including Second Language Varieties of English on High-Stakes International Tests of English Proficiency [DI]

Yi Tan

This paper presents selected results of a larger investigation into how second language (L2) English speakers with relatively low intelligibility and familiarity to the targeted L2 listeners can be used on the listening sections of high-stakes English proficiency test without compromising test designers' fundamental concerns over construct validity and fairness. To this end, I investigate whether the assistance of a 60-second exposure to an L2 speaker's accent improves comprehension of texts delivered by that speaker in the context of a simulated TOEFL iBT listening test.

Two L2 speakers (one Turkish and one Ukrainian) were selected as testing speakers. 317 Chinese undergraduate and graduate students, each completed two computer-delivered listening tests: a pre-test to gauge L2 listening proficiency, and an experimental test comprising counterbalanced sets of American English-, Ukrainian-, and Turkish-accented conversations (2) and lectures (2).

Many-Facet Rasch Measurement (MFRM) was used to compare passage- and item-level comprehension scores across L2 listeners at three proficiency levels (high, intermediate, and low) and four exposure conditions (audio-with-script, audio-only, no exposure, and baseline). At passage level, MFRM analyses showed that almost all significant biases pointed to lectures, though for high- and intermediate-level listeners, the biases were against two of the three experimental conditions for Lecture 2, whereas for low-proficiency L2 listeners, the biases were against baseline condition for Lecture 1. Item-level MFRM results aligned with those of passage level in the sense that for advanced and medium test takers, four out of seven test items demonstrating bias against one of the three experimental conditions came from Lecture 2, while for low-proficiency test takers, four out of six items showing bias against baseline condition were from Lecture 1. Implications of the findings for the design of high-stakes L2 listening test and for language tests bringing about beneficial consequence to a multilingual society will be discussed.

Yi Tan is a current PhD candidate from the Department of Applied Linguistics and ESL, Georgia State University (US). Her research interests are second language listening assessment, and particularly the interface between sociolinguistics and listening assessment.

She has publication on the journal System and presented at mainstream conferences, such as Language Testing Research Colloquium (LTRC), Language Assessment Research Conference (LARC), and American Association for Applied Linguistics (AAAL).

ALTE Information Sessions

What is ALTE and how can we help you?

Graham Seed and Kasia Wozniak

WED-26th

If you are new to ALTE, or an old friend who'd like a reminder, this session will take you on a brief overview of what ALTE and our main objectives are. It will also introduce the new ALTE Services Unit and demonstrate how the Services Unit can help you.

Graham Seed is the current ALTE Secretariat Manager, and he is responsible for the overall management of the Secretariat as well as the strategic planning, communications and partnerships.

Kasia Wozniak manages the ALTE Services Unit, overseeing the delivery of the courses, training and audit functions together with additional consultancy services.

Raising Standards in Assessment Literacy - the ALTE-Equals Course

Elaine Boyd

THU-27th

The European Profiling Grid (EPG) is an instrument which describes the competences of language teachers encouraging them to progress and assess their own professional development. One of the key competences, and one which is becoming much more

critical as assessment for learning gains traction, is a teacher's competence in assessment. With the aim of raising standards in this competence, ALTE and EAQUALS have partnered together to create an online, self-access course in assessment that aims to take teachers from levels 1.1. to 2.1 in the competences for assessment on the EPG. It will cover the key principles and issues in assessment with a strong focus on classroom assessment. The course will be available free of charge and is for all those interested in the assessment of, and for, learning. Trainee teachers, teachers, Academic Managers should all find it helpful and interesting. In this session we will look at the structure of the course and share some examples of activities, as well as showing how it will illustrate progression through the competency levels of the EPG.

Elaine Boyd has worked in English language assessment for a range of international testing organisations for the last 30 years. She has worked closely with teachers and teacher educators designing and delivering courses in assessment literacy in Europe, India and SE Asia and has published articles in this field. She is also the author of several exam coursebooks for leading international publishers. She is an Associate Tutor at University College, London. Her research interests are classroom assessment and managing learner feedback

ALTE Q-Mark: Practicalities

Helen Garside-Hornby

THU-27th

This is a brief session to introduce the Q-Mark auditing process to institutions who might like to go through such an audit in the future. The practicalities of the process will be overviewed, and information will be given about how decisions to award a Q-Mark are made.

Helen Garside-Hornby is the ALTE Audit Co-ordinator. The ALTE Auditing Unit ensures the functioning of the Q-Mark auditing services for Full Members and applicant members. This includes the implementation of the decisions taken by the Standing Committee and the QMS Working Group, coordinating the regular audits of members' examinations, organising training sessions for auditors and auditees, and overseeing relevant documentation.

ALTE Services Unit drop-in

Kasia Wozniak

THU-27th

Drop in at any time during this session to learn more about the ALTE Services Unit and how it can support you and your institution.

Kasia Wozniak manages the ALTE Services Unit, overseeing the delivery of the courses, training and audit functions together with additional consultancy services.

Paper Presentations

WEDNESDAY 26th APRIL

AIDA – AI in Digital Assessment – Lessons learnt from 3 use cases of converting AI research results into usable language assessment features [D]

Bert Wylín and Filip Vanlerberghe

Artificial Intelligence (AI) is certainly an important driver for many improvements in educational technology, including assessment technology. Both open and proprietary data sources and algorithms (like Google's BERT e.g.) are being integrated and further developed to fit into the world of digital assessments. But this integration is not a simple plug-in operation.

In a recent research project involving both academic (UGent and KU Leuven) and industrial partners, an e-assessment platform vendor was therefore included, to explicitly also research the feasibility and usability of AI applications. Complex algorithms often have a complex interface where many parameters must be

adjusted to obtain a usable result. Mastering this complexity cannot be expected from an "average" author of test items.

Three different integrations were/are being investigated and implemented:

- Automatic distractor generation for generic multiple choice questions,
- Automatic item generation: enriched gaps generation for cloze questions in language learning,
- Automatic estimation of the CEFRL vocabulary level of an answer to an open question (written or spoken), as an aid to the human correction of essay questions.

The more specific the application domain of AI, the more useful and employable the results will be. The automatic generation of distractors for multiple-choice questions for generic (as opposed to topic-specific) items yields quantitatively more, but qualitatively less useful results than domain-specific ones. The data hunger of the AI algorithms must be constantly fed to achieve better results. Final factor is the construction of a user-friendly user interface for the authors, both before and after the activation of the AI algorithm itself.

Bert Wylín has both an academic and a business profile. He worked at the KU Leuven since 1993, leading the Education Innovation Centre, for integrating EdTech in university education. In 2001, he founded a KU Leuven university spin-off, now Televic Education, developing and servicing educational technology projects and products. Today, Bert's focus is on the content side of e-assessment: how does technology best serve the didactic and pedagogical needs of content and skills measurement and assessment?

Filip Vanlerberghe is Product Manager at Televic, where he is responsible for product strategy, roadmap and innovation of assessmentQ, an advanced e-assessment platform. Filip applies 20 years of experience in software development, usability, e-learning, marketing and strategy. Previously, he was Director of Corporate Product Development in an international company. As an entrepreneur, he also owns and participates in several small businesses.

Recognising stakeholder interdependency through a language assessment literacy training framework in specific purpose language assessment [F]

Neil Bullock

Language assessment literacy (LAL), as a theoretical mechanism has been a topic of debate for over 20 years. Despite this time frame, many studies have focussed exclusively on conventional learning contexts, such as schools, colleges and universities. Whilst such debate is creditable in increasing awareness of LAL, it can be seen as insufficient across a wider range of language use contexts, such as professional, and vocational domains. This may well be the case where the real-world language use of test takers in workplace settings differs from a more traditional general language learning environment. The ideas of Target language Use and Indigenous criteria may be quite unfamiliar to those whose responsibility lies more in administrative or operational tasks rather than educational. Developing an awareness of each stakeholder role and how each one relates to language assessment is critical if any system is to be truly fit for purpose.

This presentation examines the system for assessment of language proficiency in the specific purpose workplace domain of aviation, notably for pilots and air traffic controllers. It examines various projects developed to increase awareness of LAL in this specific purpose domain and includes observations from recent research where stakeholders were able to express their opinions on critical factors relevant to their roles. Ultimately, the presentation brings such research and projects together into a potential framework for training stakeholders in LAL for specific purpose domains. It shows how such training can address emerging issues within this system for assessing LSP proficiency, including a more coherent explanation and application of the construct and how this should be assessed. It also demonstrates, through a more coherent and interdependent approach, how the quality of test instruments can be improved which, consequently, enhances fairness for all test takers.

Neil Bullock is a consultant and researcher in English teaching, test development, and teacher training. A language specialist for 18 years, he has MAs in Applied Linguistics and Contemporary French Studies, and recent work has included project-managing test development and developing a language Test Design Guidelines programme of

workshops for the aviation domain. He is currently Vice-President and General Secretary for the International Civil Aviation English Association, and current research interests include communicative competencies and online testing programmes in LSP.

Describing washback: teachers and students' voices in Jaén (Spain) [D]

Victoria Peña Jaenes

The increasing prevalence of English as a gateway to employment (Chávez Zambano, Saltos Vivas & Saltos Dueñas, 2017:761) and higher education worldwide has gone hand in hand with the demand for English language courses and accreditation exams over the last three decades in Spain. Despite the number of English language learners and the importance of English for employment and academic life, very few studies have researched the effect that accreditation exams may have on English language learning and teaching in the country.

This session is based on a broader research project, focused on washback, carried out in Jaén (Spain) with 131 students and 8 teachers. The presentation will explore the teachers' views of the more exam-focused and more general English courses they taught and the proficiency exams they were familiar with or preparing their students for. Their perceptions have been analysed and contextualised with their students' perspectives and discussed in the light of current research. Among other questions, teachers were asked about the students' characteristics and habits that could help them the most to succeed in a well-known proficiency exam. They also reflected on the effect that the accreditation exam had on their lessons and on their students. Both teachers and learners' views on the main challenges students face in the different components were analysed and compared.

The aim of this presentation is to have a better understanding of the washback that proficiency exams have in a very specific context, where these exams are well-established and where being able to prove one's language ability is crucial to improve academically and professionally.

Chávez-Zambrano, M., Saltos-Vivas, M. A. & Saltos-Dueñas, C. M. (2017). La importancia del aprendizaje y conocimiento del idioma inglés en la enseñanza superior. *Dominio de las Ciencias*, 3: 759-771. DOI: <http://dx.doi.org/10.23857/dom.cien.pocaip.2017.3.mono1.ago.759-771>.

Victoria Peña has worked for Cambridge since September 2018. Prior to that, she taught English to preschool, primary and secondary school students, as well as preparing adults for Cambridge examinations. Since joining Cambridge, she has been involved in creating content for teachers, collaborating in different assessment projects and has delivered presentations and webinars for teachers. Victoria holds a degree in Translation and Interpreting and an MA in Applied Linguistics and a PhD focusing on the washback of proficiency exams.

Standard setting across multi-mode qualifications [D]

Lauren Miller

Purpose of the study: As we look to the future of testing and the transition from traditional paper-based to on-demand computer-based assessment, it is crucial we ensure the comparability and consistency of standards across test modes. Recent years have highlighted the need for flexible assessment, that can be sat at home if need be. The days of sitting a paper-based test in a centre may soon be behind us, so we must think about what comes next. The focus of this research is a new computer-based version of an existing paper-based test. While the construct assessed in the tests is the same, the tests look very different. Most of the item types, the test blueprint, scoring, raw marks, delivery method and inherent standard setting methodologies are different.

The paper-based test is offered in exam series and has been running for several years and has a long established standard; with grade boundaries set and standard maintained using traditional awarding methodologies including classical test statistics and qualitative judgements. The paper-based test is not equated and psychometrics are not used at any point of the assessment cycle.

The computer-based tests are equated using common items and utilises item response theory statistics for test construction and standard setting.

It is absolutely vital that a pass on one test mode be equivalent to a pass on the other. This study culminated in the alignment of the standard on the new computer-based test to the existing standard of the paper-based test.

Presentation: The objective of this work was to set comparable standards, primarily relying qualitative judgements from a panel of subject experts. The presentation will take viewers through each of the steps undertaken to set a standard on the new computer-based

test that is comparable to that already established on the paper-based test.

Additional Author: Sarah Hughes, Pearson.

Lauren Miller is a Senior Researcher within the English Language Learning team. She has worked for Pearson since 2013 in a number of roles, predominately within standard setting and assessment design functions. In these roles she has designed and been responsible for setting and maintaining standards for a wide range of UK and international qualifications, using a variety of standard setting methodologies. Lauren is currently enrolled in the Postgraduate Advanced Certificate in Educational Studies: Educational Assessment through Cambridge University.

Language Tests in controlled computer rooms against those taken online with distance invigilation: some reflections [D]

Thomas Christiansen

Computer-based tests provide a fast, efficient and cost effective means of assessing students especially (see Chalhoub-Deville 2001, Suvorov and Hegelheimer 2013). Traditionally, they are administered in controlled conditions where supervision and invigilation can be conducted in much the same way as with paper-based tests.

The COVID emergency (2020-21) resulted in the sudden closure of educational institutions and a shift online of all teaching and testing. In this paper, we will examine the effects of such a change in the context of two specific university courses where a set of bespoke Moodle-hosted computer tests were adapted from a controlled and supervised context to a much freer online context using a general business communication platform (Microsoft Team).

We compare the results of almost 7,400 tests (Lexis and Use of English, Translation, Reading, Listening and Writing Tasks: the first four marked automatically, the last manually) taken in both controlled and online contexts to identify differences in results between both testing environments. The aim is to provide some hard data to ascertain how far online distance-invigilated tests taken can be adequate substitutes for more traditional formats especially in view of both their positive aspects (e.g. greater accessibility and

inclusivity for students) and their negative ones (e.g. the greater opportunity for malpractice).

The latter aspect we will examine through a close analysis of tests taken by a random sample of students in both types of test conditions in order to identify anomalies in student performance in tests over time and in answer patterns within individual tests (e.g. getting easy questions wrong and harder questions right or having correct or incorrect answers clustered at specific points in the test – see Dubner and Levitt 2009), or evidence of copying in Written Tasks.

Thomas Christiansen is an associate professor in English Language and Translation at the Università del Salento (Italy) and Director of the University Language Centre. He has taught in various positions at various universities in Apulia (Italy), the UK, and Poland. He completed his PhD in textual linguistics at Salford (UK). He has researched into various areas of linguistics including systemic linguistics and functional grammar, varieties of English, ELF, teaching English, language testing, and analysis of different corpora, including spoken discourse. He has also worked with various English language international examination boards in various capacities, including as an expert consultant.

The Test as an Opportunity. The Case of Less Widely Tested Languages [DI]

Dina Vilcu

Language tests are mainly perceived as instruments responding to the concrete need of proving language skills. This need is mainly shaped by the international world of business and education. The huge business of testing in a small number of very intensely used languages around the world mirrors this situation.

The main purpose of this presentation is to propose adding a new dimension to language testing: the test as an opportunity. This dimension should impact mainly on less widely tested/taught/used languages. It aims to go beyond the concrete activity of administering examinations, towards contributing to multilingualism, and more than that to the preservation of language diversity. Concrete actions at this level might be the creation of proficiency language tests for the less widely tested languages (LWTL), when there such an instrument does not exist and supporting the ones already created.

In this paper I intend to present concrete examples of opportunities language tests can represent for less commonly used, minority and endangered languages. These examples are related to:

- producing language learning materials and technology and contributing to the professionalisation of the practitioners in the domain;
- maintaining the native speakers' positive perspective on their own language, including when they choose to emigrate;
- broadening the adult and especially the young learners' choice of languages they learn and use; encouraging young learners to also choose less spoken languages;
- contributing to linguistic democracy and linguistic human rights, empowering less widely used, minority and endangered languages.

Languages are instruments for communication, culture and tradition keepers and personality and community builders. However, they also need to be acknowledged the role of shaping mentality. Language tests can contribute to maintaining a positive mentality on less spoken languages; we just need to project them in the right light. The present paper is intended to contribute to this purpose.

Dr Dina Vilcu teaches and assesses Romanian as a foreign language (RFL) at Babeş-Bolyai University, Cluj-Napoca. She teaches general language and LSAP courses, as well as tailor made courses for interpreters and translators. She also teaches courses on LSP and language assessment at MA level. She is interested in the challenges involved in language assessment for less widely tested languages (LWTL). Other dimensions of her research are general linguistics and sociolinguistics.

Early perspectives on high-stakes assessment conducted online; the complex relationship between test security and rating candidate performances in an at-home setting [D]

Tony Clark and Emma Bruce

If we accept the assertion that language assessment has the power to shape aspects of society, post-pandemic changes to how a globally dominant high-stakes test such as IELTS is delivered provide a solid case to investigate what this can mean in practical terms.

The newly developed secure test delivery platform introduced in 2022 for IELTS Online (IOL) promotes accessibility for test-takers previously reliant on attending test centres. Although an online testing capability may be considered a prerequisite in the contemporary assessment domain, it would be disingenuous to overlook the challenges this brings alongside the opportunities. Ensuring that IELTS remains fit for a digital future requires extensive validation research to support its migration to online delivery. This evidence should be publicly documented (Clark et al, 2022), and form part of an ongoing program of validation work that includes views of key stakeholders (Chalhoub-Deville & O'Sullivan, 2020). This is necessarily a gradual process for IOL, due to its high-stakes nature and the international scale at which it operates.

Regarding individual components, the speaking section of IOL essentially uses Videocall Speaking (VCS) which has been extensively trialled previously (see Nakatsuhara et al. 2021). Listening, reading and writing are largely the same as the existing computer-delivered test available in designated centres. What remains to be investigated is IOL as an overall online testing experience, including all aspects of secure at-home delivery and the impact of this on candidate performances and the rating process. This presentation gathers input from test-takers and proctors through a series of initial surveys and interviews. Findings indicate that the relationship between the at-home testing environment, proctoring and allowing candidates a fair opportunity to demonstrate their language proficiency is complex. Recommendations for online assessment for IELTS and the broader assessment community are included.

Tony Clark is Head of IELTS Research at Cambridge University Press and Assessment. His PhD focused on the relationship between high-stakes test scores and candidates' subsequent higher education experiences. Comparing different tests and their implications for student performance is a key research interest of his. Tony has published and presented widely in these areas, and language assessment and pedagogy research more broadly.

Dr Emma Bruce is a Researcher with the Assessment Research Group at the British Council. Emma has a specific focus on research and validation of IELTS and also works on range of projects for assessment systems designed and developed by the British Council. Emma has over 25 years of experience working in the tertiary sector, both in the UK and overseas, initially as a teacher, and later as an EAP testing and assessment specialist. Emma's main interests are in EAP and integrated assessment.

Post-entry language assessment in a diverse open access university system: the role of academic and general language ability [DI]

Jordi Heeren and Lieve De Wachter

The transition from secondary to university education is often conceived as difficult. The increasing diversity in student backgrounds poses new challenges for universities (Read, 2015). An intrinsic part of students' background is their language ability. Not only international students struggle with university language requirements, many domestic students experience difficulties as well, especially in systems with low entry requirements. Here, post-entry language assessments (PELAs) are often used to detect possible at-risk students that might benefit from remedial activities to increase their chances of study success (De Wachter et al., 2013; Kuiken & Vedder, 2021; Read, 2015; Van Dyk, 2015).

In this study, we investigate how two language proficiency tests in an academic setting relate to students' academic achievement: an academic language screening test and the ITNA computer test, part of a university entrance test in Flanders that focuses on general language proficiency. We show that both constructs are strongly related, but that academic language proficiency scores render more meaningful variance and can select a larger potential at-risk group, as almost no students fail the ITNA computer test. However, the ITNA threshold and the thresholds in the language in use subtest, do seem to detect a small, but high-risk group. Regression analyses using students' demographic and educational background variables show that only academic language proficiency remains a small significant predictor of achievement when background factors are controlled for.

In all, this study shows that in a population of mostly domestic L1 students low stakes PELAs can be useful, especially when they focus on academic language. However, general language ability might have a role to play as well, as it can point to a very small but more clearly defined at-risk group that might also benefit from additional general language instruction.

Additional Authors: *Leen Verrote, Anne-Lies Nordin and Ines Blomme, KU Leuven.*

Jordi Heeren works at the Leuven Language Institute (KU Leuven) as a lecturer on academic language skills in the faculties of Arts and Social Sciences. His research focuses on (academic) language proficiency in the transition to higher education.

Lieve De Wachter is a professor of Academic Writing at the University of Leuven (Belgium). She is head of the KU Leuven Writing Centre and her research focuses on Languages for Academic Purposes and Digital Writing Support. Together with her team she has developed the Digital Writing Aid Dutch and the TaaIVaST-test for strategic academic language skills.

Promoting positive washback to teaching and learning through understanding of language frameworks [F]

Margaret Malone and Howie Berman

By accurately assessing student progress and proficiency, test results provide positive washback to learners, teachers, and programs (Tsang & Isaacs, 2021). However, language teachers often have different understandings of the frameworks used to develop tests and thus how to review, report and use the outcomes for teaching and learning (Malone, 2013). It is critical to find ways to explain test results and the frameworks used to develop them in ways useful to learners and teachers.

This paper describes a computer-delivered test in 11 languages developed in conjunction with secondary and primary language teachers. It shows how practitioners can and do apply different aspects of the test's development, frameworks and subsequent outcomes data to adjust their instruction. Based on its test specifications, the test affords multiple opportunities for washback in terms of classroom instruction and curriculum and lesson planning. First, sample tasks allow instructors and learners to better understand: 1) the types of constructs being assessed, and 2) how to navigate the test platform and prepare for online administration. Second, learner score reports provide specific feedback on areas of strength and challenge, as well as comparisons in performance to similar groups of learners. Finally, the learner score reports, as well as broader school and national reports, allow teachers to reflect on student progress on the standards that inform the framework. Evidence will be shared for ways that teachers and world language coordinators have aligned their pedagogy with this assessment.

Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329-344. Tsang, C. L., & Isaacs, T. (2021). Hong Kong secondary students' perspectives on selecting test difficulty level and learner

washback: Effects of a graded approach to assessment. *Language Testing*, 02655322211050600.

Additional Authors: *Camelot Marshall and Reuben Vyn.*

Margaret E. Malone (Ph.D., Georgetown University) is Director of Assessment and Research at ACTFL. She has nearly three decades of experience in language test development, materials development, delivery of professional development, and teacher training through both online and face-to-face methods, data collection and survey research, and program evaluation. Her current research focuses on language assessment literacy, oral proficiency assessment, the influences of the Seal of Biliteracy on language teaching and learning, and the development of short-cut measures of proficiency.

Howie Berman is an association executive with 20 years of experience working with non-profit membership organizations, Executive Director for ACTFL, the largest U.S. organization serving the needs of language educators and learners. As ACTFL's chief staff executive, he is responsible for continuing to enhance ACTFL's reputation and influence in the language education community. He received his BA in Political Science and Spanish from the University of Richmond and his MA in Middle Eastern Studies from the University of Chicago.

Advantages and practicalities of a flipped classroom approach to standard setting [D]

Jane Lloyd

In response to the COVID-19 pandemic, a computer-mediated approach to standard setting has been developed or is under consideration by many testing organisations. This switch to a completely online approach has benefits, such as lower costs in several areas, and easier sharing of non-confidential material, instructions, and links to online surveys (Katz & Tannenbaum, 2014). The setting of this paper is an ongoing programme of alignment of English language tests to the CEFR. In particular it describes the standard setting for Linguaskill, an online multi-level test of English for adults, available online. In addition to financial benefits, the move to an online environment also improved operational practice through increased inclusion and diversity of participants. A flipped classroom approach was employed, with online meetings devoted to plenary and small group discussions. Panellists worked

on standard setting tasks between online meetings, and this resulted in a reduced cognitive load for participants, with more time available to participants for considered judgement and quiet individual reflection. This paper addresses online remote aspect of the standard setting, outlining lessons learned in the setting up and running of panels in an online, global, flipped classroom environment, provides examples of how the two popular standard setting methods (Cizek & Bunch, 2007) of modified Angoff and Bookmark can be deployed online, and how validity evidence can be gathered using commonly available software.

Jane Lloyd is a Senior Research Manager at Cambridge University Press and Assessment. Before joining Cambridge, Jane worked in Japan, Southeast Asia, Turkey and the Middle East as a teacher, trainer and manager in state schools and higher education for over 20 years. She has a degree in Linguistics, a DELTA, an MA in Linguistics and TESOL, and an MA in Language Testing. She is currently studying for a PhD in Language Testing at CRELLA.

Automated Remote Proctoring in Language Assessment: Is the Future Already Here? [D]

Larissa Leifland Berntsson

Information technologies have been increasingly important in language assessment over the past years. Moreover, the COVID-19 pandemic has forced institutions to quickly react to the changing environment of language assessment, and rapidly implement new technologies in how the testing is delivered. Together with education, assessment has largely moved online, presenting new opportunities and new challenges for language experts.

In 2020, the Saint Petersburg University designed and launched a whole new online platform for the Test of Russian as a Foreign Language (<https://languagecenter.open.spbu.ru/>). After nearly three years of piloting and practice, we are ready to report back on the method used, providing reflection on it and a performance evaluation. The paper outlines the experience of automated remote proctoring, its benefits and drawbacks, and also sets out a perspective of its subsequent development.

The platform provides an opportunity to take the exam from different locations, which inevitably raises important questions, such as fairness, reliability and security of the tests, as well as privacy and data protection for the candidates. An attempt to address these

issues was made, and a number of solutions for them are provided in the paper. Alongside with these, the data includes statistical analysis and evidence collected from surveys and interviews with examiners, test takers and other stakeholders. The findings might help other institutions to better address issues similar to those that this platform has faced in the past and might face in the future and ensure a smoother implementation process.

Additional Author: *Nadezhda Dubinina, St Petersburg University.*

Larissa Leifland Berntsson. Lecturer in Russian Language, London School of Economics and Political Science. Russian Language Assessment Expert, St Petersburg University

The Magic of Machine Learning: How to Develop Tests in Multiple Languages Simultaneously and At Scale [D]

Sarah Goodwin

Spanish and French are two of the most commonly learned additional languages worldwide (Blanco, 2021). Regardless of how learners are acquiring languages (school, online, tutoring, etc.), they may need to provide language proficiency certification for educational, occupational, or other purposes. Without formally certifying, learners describing their level must self-assess (e.g., CEFR [Council of Europe, 2001]; ACTFL [The National Standards Collaborative Board, 2017]), or estimate based on the numbers of semesters or years they have studied the language. Given the diversity of language learning contexts across the world, this self-assessment results in lack of interpretation across curricula and tests. In addition, there are fewer low-cost, accessible options for Spanish and French than for English. Thus, there is a demand for lower-stakes assessment results that harness learner data from around the world. By combining machine learning, engineering, and test development expertise, language tests and related learning and curriculum tools designed for the digital era are well positioned to tackle this challenge. The present project describes efforts toward item generation and difficulty estimation for Spanish and French assessments.

The items for the assessments are designed to assess both comprehension of audio and visual media (listening and reading

reception). The assessments use four computer-adaptive task types to assess understanding of: vocabulary in context, morphosyntax, phonology and word boundaries, and orthography. Multilingual transformer-based models (RoBERTa; Liu et al., 2019) and other machine-learning approaches were applied to estimate item difficulty even before the collection of item response data, enabling the creation of a large item bank. Data for the project included expertly rated Spanish and French vocabulary items and texts that were created for a large digital language learning platform. In addition, large word-level corpora with frequency and usage information were used. Project findings are described, including applications and implications for assessment in Spanish and French.

Additional Authors: *Lauren Bilsky, Phoebe Mulcaire and Burr Settles, Duolingo*

Dr. Sarah Goodwin is an Assessment Scientist at Duolingo, specializing in language test development. Previously, she was a postdoctoral fellow in measurement in a child literacy laboratory. Sarah has developed and scored test items for examinees from children to adults, including dual language learners of English and other languages, and people with disabilities. Her work can be found in Assessing Writing and Language Assessment Quarterly. She holds a Ph.D. in Applied Linguistics from Georgia State University.

Recommended validity periods for high-stakes language test scores: How are these understood and applied in HE, medical, and immigration sectors? [DI]

Emma Bruce, Karen Dunn and Tony Clark

Several high-stakes English proficiency tests suggest a two-year time-limit on validity for score usage (IELTS; Pearson; TOEFL iBT). Although such timeframes provide a useful rule-of-thumb for recency of testing, adherence to such limits can have far-reaching consequences for individuals and institutions. Firm limits on validity can restrict achievement of goals and staffing of crucial functions, whilst, conversely, a relaxed approach may risk over-estimation of language proficiency. In their integrated argument-based approach to validation, Chalhoub-Deville and O'Sullivan (2020) emphasise the importance of considering consequences for stakeholders,

stressing the need to tailor communication to engage all intended user groups. This is particularly relevant for test score currency and use.

In this presentation we share preliminary research which systematically collates and compares policies regarding test validity periods in five countries and three key sectors (higher education institutions, medical regulatory bodies, and immigration authorities). An analysis of policy documentation was conducted using publicly available information from 90 universities and 18 medical regulators, plus four immigration authorities. This information is considered alongside qualitative insights from in-depth interviews with stakeholders.

Findings indicate that the recommended two-year validity period is overwhelmingly the norm, with only a few exceptions. Score users place trust in test developers and tend to adopt recommendations as policy (cf. Hamid et al. 2019; O'Loughlin 2011). This lack of contextual variation taps into the "equity vs. equality" dilemma observed by Lam et al. (2021) in the HEI context, and reinforces the value of fostering shared understanding for the basis of such recommendations (Pill and Harding, 2013).

Moving forward, we plan to work towards developing practical recommendations for communicating issues around test validity periods across different contexts of use, accounting for risks, responsibility, equity, and fairness. We hope that this presentation will lead to lively discussion on the possibilities and limitations of such a goal.

Dr Emma Bruce is a member of the Assessment Research Group at the British Council and her specific focus is on research and validation of IELTS. Emma obtained her PhD from CRELLA having researched the impact of time allowances in source-based writing assessment in the EAP context. Emma has over 25 years of experience working in the tertiary sector in the UK and overseas, initially as a teacher, and later as an EAP testing and assessment specialist.

Dr Karen Dunn is a Senior Researcher in measurement and evaluation at the British Council. She holds a PhD in Applied Social Statistics and Masters in Language Studies. The focus of Karen's PhD research was on using Explanatory Item Response Theory to investigate word difficulty for L2 learners of English. In additional to operational test concerns, her current research interests include scoring validity of reading reordering tasks, assessing language test dimensionality, and linking motivational profiles to proficiency outcomes.

Dr Tony Clark is Head of IELTS Research at Cambridge University Press & Assessment, in the Research Group. Other areas of

research interest include test preparation, diagnostic assessment, and accommodations in language testing. He has published in several major language assessment and applied linguistics journals, regularly enjoys presenting and attending international conferences, and is an active member of the academic community.

Cuestiones pendientes en la evaluación de la escritura creativa [F]

Jorge J. Sánchez Iglesias

La presencia de la escritura creativa (EC) es, cuando menos, irregular en distintos frentes. De la misma manera que su espacio varía enormemente en diferentes sistemas educativos (por ejemplo, está mucho más desarrollada en el Reino Unido que en España), es fácil constatar su desigual representación en el ámbito de las lenguas segundas o extranjeras, tanto en la investigación como en la práctica. En cuanto a la primera, faltan estudios adquisicionistas (en relación con cualquier lengua), por ejemplo, sobre el itinerario de desarrollo de la EC, sobre la interacción de las subcompetencias que la conforman o sobre la contribución de dicha actividad al proceso general de adquisición de la L2. En cuanto a la segunda (centrándonos ya en el caso del español LE/L2), apenas se ha pasado de loables recopilaciones de propuestas surgidas para favorecer su inserción en el aula, aunque en muchos casos carentes de un componente reflexivo desarrollado a la hora de favorecer su evaluación. La EC disponía de una escala propia ya en el MCER (con particularidades, eso sí, en su elaboración). En el Volumen complementario se han añadido nuevos descriptores, además de sumarse otras menciones en relación con la elaboración de textos que responden a trabajos creativos (incluyendo obras literarias).

El objetivo de esta propuesta es doble. Por una parte, estudiaremos la progresión que ha experimentado la EC desde el MCER hasta el Volumen complementario, considerando de qué forma puede fomentar la expansión de la EC. Por otra, analizaremos hasta qué punto la escala disponible en el Volumen complementario puede servir como base en la elaboración de rúbricas que resulten aplicables para cualquier actividad de EC, considerando que los descriptores disponibles resultan en cierto modo limitados (por ejemplo, en términos de tipos textuales) para la diversidad de propuestas propia de la EC.

Jorge J. Sánchez Iglesias (PhD) read Spanish and Italian at the University of Salamanca (Spain). He is a lecturer of Spanish and Discourse Analysis at the School of Translation and Documentation (Library Science). He previously taught Spanish as a foreign language

at Cursos Internacionales. His research focus on assessment and reading/writing interfaces, considered from an applied perspective (first/second language learning, translation).

Assessment of existential competence of learners in language education: Issues, Frameworks, Modelling and Standards of performance [F]

Denise Lussier, Ph. D. and Isabelle Lussier, M.A.

Within the evolution of plurilingual and multicultural societies, language education enters a new era. Schools are viewed as institutions that must foster social cohesiveness, promote values and attitudes that will be accepted by diverse communities and bring students of different ethnic groups to develop positive awareness of other cultures, based on mutual respect and even empathy. In this perspective, challenges extend beyond linguistic and language competences. We need to address the issues regarding the development of intercultural and transcultural competences as current research views language teaching and learning as a discipline which embodies by nature the presence of another culture, contact with alterity and mediation in the interactions with members of the other culture.

This paper is divided in two parts. First, it summarizes a research project using multidisciplinary qualitative and quantitative approaches linked to the affective domain of language learning. The aim was to understand the construct of schemata that reflects learners intercultural and transcultural representations as positive (associated with xenophilia) or negative (linked to xenophobia). It focuses on the validation procedures carried out to measure «existential-being» competence as socio-affective factors to influence the development of CIT competences based on existing frameworks (Lussier, 1997; Byram, 1997 & CEFRL, 2001). Exploratory analyses confirm the validation of the structure of a conceptual framework and the modelling of such framework (Lussier & al, 2010; 2014) which should allow reliable coherence in curriculum, learning outcomes and assessment developments. Second, it presents guidelines to assess CIT existential-being competences, three levels of complexity and socio-affective profiles that can be used as standards of performance.

Denise Lussier received her Ph. D. in educational psychology and language assessment from Laval University and a M.A. on Curriculum

Development from Boston University. She is currently Emeritus Professor and researcher at McGill University. Her research projects include the development of Language Benchmarks and Assessment instruments in English as professional language, for Quebec Health Professionals (2010-2021). She also co-coordinated projects for the Council of Europe: Cultural mediation in language learning and Guidelines for the assessment of intercultural communicative competence. E-mail address: denise.lussier@mcgill.ca.

Isabelle Lussier holds a Master of Arts in Measurement and Evaluation (UQAM), a Bachelor of Education in TESL/TFLS from McGill University and a Certificate in Psychoeducation from the University of Montreal. She is currently Director of Studies at Collège Charlemagne. She presented workshops on the Validation of a survey questionnaire to explore the concepts of Xenophilia/Xenophobia (UQAM, 2001), The results of the survey questionnaire (ADMEE, 2004) and the Validation of a conceptual framework for intercultural communicative competence (AERA, 2005). E-mail address: lussierdocimologue@hotmail.com.

Killing a flock of standard-setting judgements with one digital stone [D]

Anniken Telnes Iversen, Rønnaug Katharina Totland and Torbjørn Ohlseth

Standard setting the listening and reading tests of the official Norwegian test for adult immigrants is a daunting task: In a multi-level, partly adaptive test with four levels and two skills, there are eight cut scores to set, from A1 to B2. An ideal standard-setting process might take several weeks, with thorough training, familiarisation, and workshops, but would be costly in time and money. Norway is a small language community, and we need to supplement in-house testing expertise with teachers who know the CEFR and candidate group. However, getting teachers 4-5 days release from teaching was regarded as unrealistic.

Exploiting digital developments from the Corona pandemic, in spring 2022 we conducted standard setting via online meetings, which enabled us to include teachers from around Norway. Flexible learning ideas made training and familiarisation more flexible and pedagogical: We offered short films and tasks to do at home, in addition to regular session before the group discussions.

Group deliberations were conducted in breakout rooms and sharing screens for discussing tasks and adjusting judgments.

The standard-setting method was the same as in former standard settings, the “data-driven direct consensus method” (3DC) developed by CITO in the Netherlands. Although judges were more diverse than before, and many had no experience of standard setting, the standard error and cut scores were not much different.

What was very different, however, was that internal judges experienced much less time pressure, group deliberations improved, teachers offered valuable insights, and their inclusion gave the process more legitimacy. Feedback from judges was also very positive, and for teachers the sessions functioned as further education in language testing and gave them a better understanding of the tests.

Anniken Telnes Iversen, Rønnaug Katharina Totland, and Torbjørn Ohlseth are test developers, who have worked particularly with the reading and listening test of Norwegian for many years. Rønnaug Totland has been part of the development from a paper and pencil test to the online test we have today.

Online Proctoring of Language Examinations: A Survey of Candidates' Attitudes and Perceptions. Languagecert online ESOL EXAMS and USAL esPro test [D]

Marta García, Yiannis Papargyris and Leda Lampropoulou

Assessment has long been conducted in front of an invigilator in a test centre such as a school or university, or in a computer laboratory under examination conditions. Alternative modes of test administration – in addition to the traditional setting described above – involve a test taker taking a pen-and-paper (or a computer-based) test in their own home, invigilated by an examiner (proctor) in another location (via video link) or a speaking test administered remotely by an examiner. Taking an exam unsupervised, as is the case with take-home exams, is also sometimes the case.

EsPro has been using PeopleCert's online proctoring system for several years, now. In this presentation, we will have the opportunity to present PeopleCert's proprietary online-proctoring system, its security features as well as its impact on the efficiency of the exam delivery process altogether. Furthermore, we will be discussing the findings of a recent survey of candidates' attitudes and experiences with online exams.

Marta García. Spanish as FL teacher in Cursos Internacionales de la Universidad de Salamanca and responsible for the LC USAL esPro test, holds a B.A. 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.

Paper Presentations

THURSDAY 27th APRIL

Language tests for Citizenship: A Review [DI]

Miguel Fernández Álvarez

Traditionally, citizenship tests (including the linguistic component) have been delivered as paper and pencil tests, and test candidates were required to do the exam at a testing center for security reasons. However, the current fast-paced digital world we are living in is starting to have an influence on the way languages for citizenship purposes are being assessed. Some tests have already been transformed into computer-based tests that candidates have to take at a testing center (Cooke, 2009), while others can be completed from any location. Web based exams have become generalized in the last twenty years, but there has been a special need for online testing worldwide during the 2020 crisis due to the COVID-19 pandemic, which will influence the way languages will be assessed.

This presentation focuses on how technology is shaping language tests for citizenship in order to respond to the new demands of stakeholders and to adjust to the new realities we are facing in the present era. An analysis of citizenship tests (with a focus on both the civics and the language components) is offered in the first section of the paper. After this initial overview, our attention turns into computer-based language citizenship tests and some of the

issues associated with them, such as proctoring and security. In this section we introduce and describe the concept of e-securitization, which brings together the fields of language testing, cyber security and artificial intelligence. Before offering our conclusions and a brief overview of future areas for research, we present some examples of computer-based language for citizenship purposes used in several countries together with apps and resources available to prepare for such tests, including examples of Massive Online Open Courses (MOOCs).

Additional Authors: *Jesús García Laborda.*

Miguel Fernández is an Associate Professor in the Department of Linguistics Applied to Science and Technology at the Universidad Politécnica de Madrid. He holds a Ph.D. in English Philology (University of Granada, Spain) and two Masters degrees: M.A. in English Philology (University of Granada, Spain), and M.A. in Language Testing (Lancaster University, UK). His areas of interest include Bilingual Education, English for Specific Purposes and Language Testing.

Application of the CEFR Companion Volume to a cognitive view on pragmatic competence [F]

Cristina Heras Ramírez

Learning a language goes beyond the knowledge of linguistic forms; it also involves developing the ability to appropriately and purposely function in another language. This ability is captured by the CEFR's pragmatic competence, one of the communicative language competence, which encompasses discourse, functional, and design competence. Nonetheless, the descriptors have been criticized for being imprecise leaving the implementation of CEFR in hands of individual practitioners (Alderson, 2007; Eizaga-Rebollar and Heras-Ramírez, 2020; Figueras, 2012; Little, 2006; North, 2014; Sickinger and Schneider, 2014). Operationalizing a construct of pragmatic competence is indeed challenging since it involves the gradual development of complex cognitive processes. To capture the developing stages, Ifantidou (2014) approached pragmatic competence from a relevance-theoretic stance as an information processing ability that involves different skills, namely, pragmatic awareness and metapragmatic awareness. My aim is to redefine pragmatic competence from a cognitive view to improve the CEFR's approach. Based on the critical analysis of the illustrative scales,

the present paper examines to what extent the CEFR incorporates pragmatic and metapragmatic awareness aspects. Implementing a new definition of pragmatic competence would enhance the guidelines to ensure a more objective application of the CEFR on examinations, curricula, or teaching material.

Cristina Heras Ramírez is a predoctoral research fellow at the University of Cádiz where she is currently pursuing her PhD on Linguistics. She also holds a Master of Arts in the Teaching of Foreign Languages (MATL) from the University of Southern Mississippi and a bachelor degree on English Studies from the University of Cadiz. Her areas of interest include Applied Linguistics, Pragmatics, and Second Language Teaching and Testing.

Improving proficiency descriptors in the 21st century: Incorporating diversity, equity and inclusion [DI]

Margaret Malone and Howie Berman

Developing language proficiency descriptors is challenging, because developers must ensure that the descriptors show a logical and research-based progression both vertically (across levels) and horizontally (within levels and across sublevels) of the scale (Berger, 2020). Revising existing descriptors must not only focus on clarity of language within the scales but also on incorporation of standards for diversity, equity and inclusion (DEI) with attention to bias and sensitivity (Grapin & Lee, 2022).

This paper provides an overview of a process to review and make suggestions for revision to an internationally recognized language proficiency scale used across multiple languages, contexts, age groups and countries, as well as the considerations used in both the review process and change-making procedures. First, the paper outlines the origins and use of the scale in language teaching, learning, and testing. Next, the presenters describe the iterative process of reviewing the scale for fidelity to research, language teacher preparation, ongoing development in language teaching, learning and assessment, consistency with its current use, and language and ideas expressed in the scale. Multiple approaches were used to gather input, including internal and external expert reviews, focus groups, and iterative discussions. The paper also describes a multi-phase approach to reviewing the proposed changes to descriptors both vertically and horizontally for consistency. Most importantly, the presentation focuses on the emphasis of DEI to increase accessibility of the scale.

Berger, A. (2020). Specifying progression in academic speaking: A keyword analysis of CEFR-based proficiency descriptors. *Language Assessment Quarterly*, 17(1), 85-99. Grapin, S. E., & Lee, O. (2022). WIDA English language development standards framework, 2020 edition: Key shifts and emerging tensions. *TESOL Quarterly*, 56(2), 827-839.

Margaret E. Malone (Ph.D., Georgetown University) is Director of Assessment and Research at ACTFL. She has nearly three decades of experience in language test development, materials development, delivery of professional development, and teacher training through both online and face-to-face methods, data collection and survey research, and program evaluation. Her current research focuses on language assessment literacy, oral proficiency assessment, the influences of the Seal of Biliteracy on language teaching and learning, and the development of short-cut measures of proficiency.

Howie Berman is an association executive with 20 years of experience working with non-profit membership organizations, Executive Director for ACTFL, the largest U.S. organization serving the needs of language educators and learners. As ACTFL's chief staff executive, he is responsible for continuing to enhance ACTFL's reputation and influence in the language education community. He received his BA in Political Science and Spanish from the University of Richmond and his MA in Middle Eastern Studies from the University of Chicago.

Investigating potential bias in testing migrants' language proficiency in Switzerland [DI]

Hrisztalina Hrisztova-Gotthardt

In 2018, language proficiency requirements became part of the Swiss national naturalization law. Certain language requirements were also introduced for temporary and permanent residency. In this context, an obvious need emerged for a reliable tool to assess the migrants' language competences in three of Switzerland's official languages, namely French, Italian, and German. In response to this need, the fide test was commissioned by the State Secretariat for Migration. The test was aimed to assess the (language) ability of migrants to master everyday communication in Switzerland. Prior to defining the test construct, a needs analysis was carried out to identify the most common real-life situations in which migrants and Swiss residents interact, such as renewing

a passport, booking an appointment with a doctor, using public transport, etc. These frequent contact situations formed the basis for the communicative tasks, which candidates must accomplish during the fide test. Since the vast majority of migrants living in Switzerland are, most probably, familiar with the communicative situations on which the fide test tasks are based, it is assumed that the test reflects the same construct for all test takers and does not advantage or disadvantage any individuals or test taker groups. To check this hypothesis the results of the French and German tests completed between January and July 2022 are statistically analysed and checked for potential systematic differences across candidate groups. The analysis is performed with the goal of examining whether – with regard to the overall test results - there is any difference (i) between male and female test takers and (ii) across test taker groups with different first languages.

Additional Author: *Gábor Szabó, Department of English Applied Linguistics at the University of Pécs, Hungary.*

Hrisztalina Hrisztova-Gotthardt received her MA in German Studies at the University of Sofia, Bulgaria. In 2009, she obtained her PhD in the field of Applied Linguistics from the University of Pécs, Hungary. Hrisztalina has been acting as a ECL test developer and test coordinator since 2008. She is currently employed as a chief quality officer at the Secretariat fide (Bern, Switzerland). Besides language assessment, her research interests include paremiology and paremiography, and verbal and visual humour.

Citizenship tests as a means of inclusion. How far have we gone till now? [DI]

Thomas Rousoulioti and Maria Iakovou

The process of citizenship concerns multiple criteria which touch upon language, cultural, historical, geographical and political knowledge of the host country that the applicant shall fulfil. The main naturalization processes can be different among countries (like some EU states and the U.S.A.). Emphasis is placed on the citizenship exams that take place in Greece. The main goal of this presentation is to highlight the points that have been amended within the former framework for the citizenship exams, as the Greek citizenship exams were not thoroughly standardized (Kokkinidou et al., 2017). From 2020 and on (Law 4735/2020) the exam process for the acquisition of citizenship in Greece has changed (<https://>

www.exetaseis-ithageneia.ypes.gr). This presentation attempts to shed light on what and how has changed as well as what still needs to be done regarding the citizenship exams, as it still remains duty of the host country to respect the rights of the migrants test takers. Finally, this presentation discusses useful lessons that might be drawn from best practices in other countries. Acquiring the Greek citizenship de facto means the acquisition of the EU citizenship for migrants in Greece. Under these conditions it becomes clear that these are high-stakes exams in terms of the lives of migrants and the examination process should be fair, accurate, reliable and most of all fit for purpose (Alte, Lami group 2016).

Additional Author: *Kokkinidou Anna, Democritus University of Thrace*

Thomas Rousoulioti, PhD in Didactics, works as a teaching fellow at the Aristotle University of Thessaloniki, Greece. From 2010 till 2017 she worked at the research Center 'Centre for the Greek Language' where she was involved in the design of research programs for the teaching and the assessment of Greek as L2. She is a member of the steering team of SIG for inclusive assessment in Education (AEA Europe), of EALTA, ALTE and the Research Team EnA of OsloMet.

Maria Iakovou is an Associate Professor of Applied Linguistics, specializing in the teaching of Greek as a second/ foreign language, at the Department of Linguistics, Faculty of Philology, National and Kapodistrian University of Athens (NKUA). She studied Linguistics at the NKUA and the University of Reading. Her main research areas cover both the learning and teaching of Greek as L2 to migrants and refugees. Her main goal is to promote a significant relationship between practice, language theory/research and Teacher Education.

Expanding frameworks: New developments in vocational language testing [F]

Magdalena Igiel

Vocational language courses in Germany are building on integration courses and prepare migrants for the working environment in Germany. The aim of vocational language courses is to expand the already attained general language competence of learners and to increase their chances on the labour market. The achievement of a higher CEFR level and the expansion of language competence from general to workplace-related language use are the two targeted dimensions in these courses.

This paper provides background information about the extensive preparatory work on the description of learning objectives, which apply equally to course and examinations and contribute significantly to a coherent course system in the sense of constructive alignment. The development of examination formats was based on these learning objectives. The fields of action and learning objectives included in the catalogue were developed independently of CEFR levels, because the working environment contains language requirements that are independent of the respective German language skills. However, the subsequent specific learning objectives were broken down into so-called language actions reflecting CEFR levels A2-C1.

Thus, developing a framework with a new set of descriptors, which are then used to create an item bank offers another perspective on how frameworks can be quite adaptable while acting as a reference. At the same time, the constructive alignment implemented here ensures the quality of the vocational language courses.

Based on the corresponding steps of validation, it can already be stated that the examinations have experienced a high level of acceptance and that they are assessed by the vast majority of those involved as a valid form of final competence measurement in the vocational language course.

Magdalena Igiel is a test developer and digital learning specialist at telc gGmbH. She managed the new vocational language test suite development. Currently, she is responsible for the digital assessment process and quality review. She holds a PhD in International Communication from the University of Oklahoma; her research focused on the semiotics of social change in international contexts. Before joining telc, she taught a variety of in-person, hybrid, and online communication courses at the New York City College of Technology.

CEFR-J x 28 multilingual can do test development: an interim report [F]

Yukio Tono and Masashi Negishi

In this talk, we will give an interim report on the development of CEFR-J based can do tests for 28 foreign languages taught as a major at Tokyo University of Foreign Studies (TUFS), Japan. The test batteries were originally designed based on the CEFR-J, the adaptation of the CEFR for localized contexts of English language teaching in Japan. The CEFR-J as a framework consists of 100 original scaled illustrative descriptors covering 5 modes of

communication (listening, reading, spoken production, spoken interaction, and writing) across 10 levels (Pre-A1 through B2.2). For each can do descriptor, a test task was developed to assess whether the user attains the given CEFR level for a specific mode of communication.

This English test battery was used to create multilingual versions of the tests. Through the translation of the English test battery, test developers in each language familiarised themselves with test designs and specifications. After creating the translated versions, they went on to create two more original test sets. So far, three test batteries (100 can-do test items each) have been created for 25 languages. We will report on the test development process and the preliminary results of the analysis of the Reading part across multiple languages (e.g. examining whether the difficulty of test items is retained across languages).

Yukio Tono is a professor in corpus linguistics at Tokyo University of Foreign Studies. He is also the Director of the World Language Center (Lingua), TUFS. He obtained his PhD from Lancaster University, UK. His main research interest includes corpus applications in language learning, especially within the framework of the CEFR. He and Masashi Negishi are the PIs of the CEFR-J project.

Masashi Negishi is a professor in language testing at Tokyo University of Foreign Studies. He obtained his PhD from the University of Reading, UK. He serves as the Director of the Lingua Test Center, where he supervises the development of language tests for various stakeholders. He and Yukio Tono are the PIs of the CEFR-J project.

A Descriptive Overview of Pretesting Assessment within the CEFRL [F]

Matthew James and Maria Isabel Requena de Lamo

With the permission of Cambridge Assessment English given the strict, confidential nature of the Cambridge Pretesting programme, we, as Foridiomas, aim to introduce a new area of research (specifically focused on Cambridge Pretesting) that analyses how a unique form of language assessment known as “pretesting” can strengthen the L2 learning process through various classroom approaches. Moreover, we conduct a comparative analysis of the Cambridge Pretesting programme against other existing, pretesting assessment programmes while analysing various assessment models that teachers can use in the classroom.

Since level tests measure language abilities before starting a course while practice tests can be used during classroom instruction, given as homework, and/or administered as complimentary sessions, pretesting can be used as a practical, developmental application by teachers in collaboration with test developers to better gauge student preparedness in Listening, Writing, Reading, and Speaking. Through proper pretesting implementation that will allow for greater interaction between students, teachers, professionals, and experts, students will be exposed to the current content and design of exams that will be used in the near future. When properly incorporated into the curriculum, exam preparedness and readiness at current levels during the course can be better understood to properly advise the student.

Hence, teachers can evolve into reflective practitioners to best guide students towards becoming autonomous learners. A more-defined profile of the student can then be constructed by creating a pre-assessment analysis of their respective progress as the contents of a pretesting feedback form are taken into account to provide scaffolding and guidance. Since students will be more confident and engaging with their own progress given the increased feedback regarding their strengths and weaknesses, test developers will be continuously reevaluating their exams based on real pre-assessment results which will then allow institutions to foster their pedagogical practices and accountability in educational programmes.

Additional Author: Dr. Tuğba Elif Toprak Yildiz.

Matthew James is Founder/Director of Foridiomas, an international research group in Madrid, Spain dedicated to linguistics, multilingualism, and STEM English. With 18 years of international academic and professional experience in English-language development and economics in the United States, Europe, and Asia, as a TEFL-certified, Cambridge Speaking Examiner, he has created a linguistics, ACES-Impact methodology that seeks to maximize the underlying economic value of international R&D in English by increasing the competitive advantage of languages and how they interact.

Maria Isabel Requena de Lamo is Director of Shakespeare Idioms, a Cambridge Speaking Examiner, Pedagogical Reviewer at Burlington Books, and Co-Founder/Director of Multilingual Events at Foridiomas. She has presented at numerous national and international conferences over the years in Spain with Enseñanza Bilingüe, TESOL-Spain, and EDUCA regarding multilingualism and language learning. She has degrees in English Language & Literature and Aptitude for Teaching from the University of Murcia and the University of Extremadura.

A Multiple Group IRT Approach to Address Diversity and Inclusion in a Reading Assessment [DI]

Ismail Cuhadar and Burcu Parlak

Test-takers are typically treated as a single group during ability estimation working on the assumption that they come from the same population. Homogeneity cannot simply be assumed in the presence of linguistic minorities in the group. Group membership might influence ability distribution and measurement model may not be invariant across groups. Violation of measurement invariance is a menace to fairness (American Educational Research Association et al., 2014). Group heterogeneity should be considered during ability estimation to preserve the diversity of response patterns (Oliveri et al., 2016).

The purpose of this study is to explore the degree of change in reading ability estimates for a linguistic minority group when within-group heterogeneity is acknowledged. Single and multiple group item response theory (IRT) were employed to compare ability estimates between groups across calibrations. Multiple group IRT supports estimating item parameters and ability concertedly for each subpopulation and is used to address measurement invariance (Bock & Zimowski, 1997). A Turkish reading assessment with 18 items was used in the study. The assessment was administered to about 9500 students with 200 students speaking Turkish as a second language (L2) and belonging to a minority language group. L2 group was treated as focal group and L1 students represented the reference group in multiple group IRT calibration. Group membership was overlooked in single-group IRT calibration. Generalized Partial Credit Model (Muraki, 1992) was implemented in both calibrations.

Results reveal that L2 group's ability estimates were lower when they subsumed into L1 group in the single-group IRT compared to multiple group IRT where group membership was considered. By contrast, L1 group's ability estimates were practically the same. The study has important implications to enhance the inclusion of minority students. Considering the presence of different populations is necessary to estimate ability correctly, ensure valid interpretations, and create equal opportunities.

References: American Educational Research Association [AERA], National Council on Measurement in Education [NCME], American Psychological Association [APA]. (2014). Standards for educational and psychological testing. American Educational Research Association. Bock, R. D., & Zimowski, M. F. (1997). Multiple group

IRT. In W. J. van der Linden & R. K. Hambleton (Eds.), Handbook of modern item response theory (pp. 433-448). Springer. Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159-176. Oliveri, M. E., Ercikan, K., Lyons-Thomas, J., & Holtzman S. (2016). Analyzing fairness among linguistic minority populations using a latent class differential item functioning approach. *Applied Measurement in Education*, 29(1), 17-29.

Additional Author: *Meltem Yumsek Akbaba.*

Ismail Cuhadar works as a psychometrician for the Ministry of National Education in Türkiye. He obtained his PhD degree from the Florida State University in measurement and statistics. His research interests focus on large-scale assessment, item response theory and factor analysis.

Burcu Parlak is a math teacher, and she works as a psychometrician for the Ministry of National Education in Türkiye. She obtained her PhD degree from Hacettepe University in math education. Her scholarly interests focus on large-scale assessment, international assessments, and diagnostic measurement.

Language tests for migrants: How to deal with stakeholders' expectations? [DI]

Sibylle Plassmann

It is always a challenge to meet all or at least most stakeholders' expectations about language tests, but especially so in the context of migration. No other area in assessment receives a comparable degree of political attention, placing tests and certificates far beyond their original educational context. Test providers must consider the multi-dimensional role their tests are playing and consider ethical implications.

In this presentation, I will focus on stakeholders' expectations and needs. The various stakeholders of assessment for migrants represent a wide range of perspectives: while test-takers need a door-opener, the authorities want a legally sound documentation of language competence, and politicians might try to meet the demand for skilled workers. We as test providers must consider these and many more demands. For taking the responsibility to provide valid assessment in this highly politicised field we need much more than the classical test developer's toolkit. We must be able to mediate

different perspectives including our own and to reach a consensus on what a test can and should achieve in a specific context.

After considering these general challenges, I will illustrate some of the possible stakeholders' expectations and ways to meet them using the German Test for Work (Deutsch-Test für den Beruf). This suite of four tests A2-C1 is part of Germany's federal programme for the integration of migrants into the labour market. This means that government and employment agencies make a strong impact on the test, which is not necessarily based on educational considerations. Language schools and teachers have their own perspective, and test-takers still another. It was a challenge to meet any reasonable expectations, to deflect others and to bring everything together in a practicable and valid test format and a meaningful communication of results.

Sibylle Plassmann is the Head of Test Development for telc – language tests. She has a background in linguistics and teaching German as a foreign language. As a test developer, she has worked on a wide range of projects, including language tests in German and English for various target groups, including the German tests for migrants, tests for medical German, and for general occupational purposes.

Embedded meaning in Canada's English-language proficiency requirements for immigration, asylum and resettlement, and citizenship [DI]

Melissa McLeod

Canada is facing a critical fast-growing need to drive economic growth while countering its ageing population and has set ambitious immigration targets for 2022-2024 of 1.3 million new permanent residents. Canada's immigration programs currently accept test score reports from only two tests – the Canadian English Language Proficiency Index Program (CELPIP) Test and IELTS General. Although English-language proficiency testing for global migration is now a common practice, much of the research to date in our field has been conceptual or limited to describing policy decisions and local contexts. In the Canadian migration context, CELPIP research to date has been mostly quantitative and focused solely on its psychometric properties. This doctoral study examines the embedded meaning in Canada's English-language proficiency requirements for immigration, asylum and resettlement,

and citizenship by exploring 1) the factors that contribute to the use of English-language proficiency requirements and their purported connection to newcomers' integration in Canada and other traditional migrant destination countries; 2) the appropriateness of CELPIP for use within the Canadian immigration context; and 3) test takers' experiences using a qualitative survey and semi-structured interviews. Pulling the three sources of data together, this paper session will focus on the barriers the author faced when trying to engage with government policy makers, areas where language testing professionals can develop policy literacy, and suggestions for both advocacy for test takers and further research.

Melissa McLeod is a PhD candidate in the Faculty of Education at Queen's University. She taught ESL/EFL for 17 years in Canada and abroad. Her primary research interests are refugee education and refugee integration.

Issues in automated scoring of language performance tests: Methodologies, practices, and ethics [D]

Jing Xu

With the rapid advancement of machine learning technologies, we have seen an increasing use of automated scoring algorithms in large-scale educational assessment, including language assessment. This has prompted professional organisations to introduce guidelines for examining the validity of automated scoring in the latest standards for testing (AERA, APA, & NCME, 2014; ITC & ATP, in press). Language testers' general reactions to automated scoring technologies have been described as "cautiously optimistic" (Xi, 2021, p. 514). While automated scoring is seen as a promising solution to the growing demand for more efficient, affordable, reliable, and learning-oriented assessment (e.g., Chapelle & Chung, 2010; Jones & Saville, 2016; Xi, 2021), scepticism, concern and criticism have also been voiced, particularly on the use of automated scoring technologies in performance-based language assessment (e.g., Chun, 2006, 2008; Fan, 2014; Xu, 2015; Xi et al., 2016; Wagner, 2020).

This presentation discusses some key issues in automated scoring of constructed responses, including training and evaluating an automarker, malpractice detection, hybrid marking, interpretability and generalisability of automated scores, and ethical considerations in using scoring algorithms. It calls for more transparency in describing the general characteristics of scoring algorithms and

the training data, more consistency among test developers in presenting validity and reliability evidence for automated scores, and a collaborative effort between language testers and computational linguists to improve automated scoring literacy and interpretability. In the end, I will provide a list of questions for test users to consider when they need to make judgment of the suitability of automated language tests for their own assessment contexts.

Jing Xu is Head of Propositions Research at the English Research Group of Cambridge University Press & Assessment in the United Kingdom. His research areas include technology-enhanced language assessment, computer-assisted language learning, and validity issues in automated scoring and feedback. He led the validation of Linguaskill, an AI-powered English language test, and is currently the research lead on Cambridge's integrated learning and assessment products. Jing is an editorial board member of the Language Assessment Quarterly journal.

Connecting an Academic English Integrated Writing Task to CEFR [F]

Mohammed Sameer

One of the key reasons to consider using integrated assessments in the EAP context is they mirror university tasks and promote the development of relevant skills for academic success. Cumming (2013) uses the word “realistic” to describe integrated tasks and Chan and Yamashita (2022) suggest these tasks develop essential skills for university settings. The Centre for English Teaching at the University of Sydney has developed an integrated writing task where learners engage with content-responsible multimodal input texts to develop a sustained position on a topic with the aim of producing a written output. Given that these types of integrated tasks require learners to utilise a combination of skills simultaneously, the identification of a construct that clearly defines what the task entails and measures is imperative. This would be most useful for developing marking rubrics to ensure accurate assessment of tasks. These rubrics are often mapped to world standard language proficiency frameworks such as CEFR to ensure the rubrics are in par with world standards.

This presentation will describe the Centre for English Teaching's integrated writing assessment task developed for academic English students intending to pursue further studies at the University of Sydney. A key focus of the presentation will be to identify the integrated writing task construct and highlight pertinent CEFR descriptors from the Companion Volume that are relevant to the task construct.

Mohammed Sameer has an M.A. in Linguistics and a PhD in Education. He teaches academic English at the Centre for English Teaching at the University of Sydney. He is also part of the Centre's Assessment Quality Team, primarily responsible for writing assessment tasks and maintaining rater standards through processes of standardisation and moderation. He is a speaking examiner for IELTS.

Evaluating an accommodations policy for candidates in a large-scale English proficiency test [DI]

Johanna Motteram and Richard Spiby

Inclusivity and the provision of equitable access to the test construct are important aspects of fairness in language assessment (ALTE, 2020). Equal opportunity and fair treatment in tests is often associated with standardisation. However, even in standardised assessments, extra support, in the form of test accommodations, is sometimes needed to allow access to test content for all.

This presentation describes a study of the accommodations policy of the Workplace Literacy and Numeracy (WPLN) Assessment, a localised, computer-delivered language proficiency test delivered in Singapore. The WPLN itself plays an important role in providing access to opportunity and the promotion of Singapore as a meritocratic society. It gives opportunities to candidates who may have had limited access to high school assessments and may be outside the educational mainstream, including students with specific learning difficulties. For these students, achievement in the WPLN provides an alternative pathway to further training and employment.

In 2020, the British Council began delivering WPLN assessments, with more than 500 candidates receiving accommodation in line with the WPLN accommodations policy, based on global exam policies for equality, diversity and inclusion. This study investigated the appropriacy, effectiveness and impact of the accommodations delivered to WPLN candidates in order to evaluate operational elements of the accommodations policy and determine whether it is fit for purpose within the context. Operational and interview data from key stakeholders in the accommodations process were gathered in order to gain insights into how the policy was implemented and what adaptation may be necessary for local conditions. The findings are discussed with reference to the literature, current practice and the changing inclusivity landscape in Singapore. Recommendations are made for improvement in the

localisation of special accommodations and operations linked to policy development, including policy dissemination to safeguard access and inclusion in large-scale test systems.

Additional Author: Gemma Bellhouse, British Council.

Johanna Motteram is a member of the Global Assessments team in the British Council. She works with government and institutional stakeholders to identify, develop, and deliver solutions to language assessment problems in education systems. She also has extensive experience in teaching English and Applied Linguistics. Her research interests include the impact of the introduction of new assessments into education systems, and what policy makers and test providers can do to enhance or mitigate that impact.

Richard Spiby works with the Assessment Research Group on the development and validation of British Council tests. His main responsibilities include overseeing analysis and investigating skills components for new and existing tests. He also works on a variety of assessment development and training projects worldwide. His areas of interest include language assessment literacy, test-taker strategies, test analysis and inclusivity in language assessment.

In-class dynamic assessment and curriculum design framework based on needs analysis for volunteer-led adult migrant literacy program [DI]

Haeng-A Kim

As a volunteer tutor for the Adult Migrant English Program (AMEP) in Australia myself, I have been invited to the tutor training workshops to share my insights on what to teach (materials) and how to teach (delivery), yet I find it more urgent to reform the program fundamentally because it does not have a fixed curriculum that tutors could use. So, the major purpose of this research project is for providing a framework to the volunteer tutors, whose native language is English, yet do not have previous teaching qualifications or experience, so that they could autonomously conduct their one-on-one tutoring based on their student's learning needs.

The research procedure includes three parts: needs analysis, designing a framework, and validating the framework. For needs analysis, I conduct semi-structured interviews with tutors and learners, classroom observations, and tutor workshops, of which

results are gathered to figure out what tutors and learners need for a better teaching and learning practice. After that, I design a framework that contains three sections: guideline for what to assess and how to assess (assessment), guideline for how to interpret its results (assessment literacy skills), and guideline for what to teach during a class with the test results (curriculum). Once the framework is developed, trials are conducted to validate the framework.

The implication of this research is that helping volunteer tutors equipped with assessment literacy skills can build the virtuous cycle of Assessment-Curriculum-Teaching and Learning-Assessment.

* Note that this research project is my master's dissertation topic, and the research is conducted with the help of the Sunshine Coast Council in Australia where the research results will be implemented.

Haeng-A Kim obtained her TESOL certificate from Georgetown University and professional postgraduate degree in Educational Assessment from University of Cambridge where she is currently studying for her master's degree in English Language Assessment. She won the TOEFL scholarship hosted by IATEFL based on her teaching experience and best-selling books in Korea and presented at the panel discussion hosted by TOEFL Global. Currently, she works for Queensland Curriculum and Assessment Authority as an assessor and volunteers at Sunshine Coast Council in Australia.

A sociolinguistic profile of the migrants applying for Portuguese citizenship [DI]

Nélia Alexandre, Paula Simões and Ana Espírito Santo

When “tests are used for migration purposes, they should meet the highest standards of quality and fairness” (Rocca et al. 2019). In order to keep language assessment tests ethically and socially sound, as well as valid and reliable, the goal of this talk is twofold: to identify the test-takers strengths and weaknesses (Egbert 2017) and, to sketch a sociolinguistic profile of the migrants applying for Portuguese citizenship.

To achieve this, we analyse a cross-section of data extracted from the tests of CAPLE, the Portuguese organisation which certifies language skills in PFL for different purposes (most wanted for citizenship purposes), applied in 2021. The corpus analysed comprises 600 written productions of the “letter” and

“invitation” genres, corresponding to a total of 66.000 words. 19 native languages were considered (those with 10 or more applicants each). The test-takers’ difficulties were manually coded in a two-step analysis consisting of (i) the identification of deviant uses with the target structures and (ii) the categorisation of these deviant uses, in particular in the domains of vocabulary (cognates, language shift, hybrids, inadequate use – “...é o melhor para almoço[target: almoçar!]”), orthography (comecou[target: começou]), and morphosyntax (number and gender agreement, use of articles, prepositions, and subject and object pronouns, tense and mood inflection, use of connectors and relative and interrogative pronouns – “...Acho __[target: que] o café dentro do museu é...”). Sociological variables, such as age, profession, and hours of study in Portuguese were considered as well.

The test-takers show a high error rate in morphosyntax, then in grapheme-phoneme correspondence and vocabulary, which is the component with lower error rates (in line with the results of Gramacho et al. 2019). Despite CAPLE tests being subject to appropriate quality control, this profiling will contribute to a better knowledge of the target audience, redesigning the tests’ content, and improving validity.

Additional Author: Abdelkarim Diane, CAPLE.

Nélia Alexandre is an Assistant Professor at the School of Arts and Humanities, University of Lisbon; Head of the Centre for the Assessment and Certification of Portuguese as a Foreign Language (CAPLE); Pedagogical Coordinator of the online course O Meu Português. She obtained her PhD from the University of Lisbon in General Linguistics (syntax of Cape Verdean Creole). Her research has focused on the following areas of Linguistics: (comparative) syntax, language contact, varieties of Portuguese in Africa, and acquisition of L2 Portuguese.

Paula Simões is the Assessment Director at the Instituto de Avaliação Educativa, I.P. (IAVE); since September 2019, she has also been deputy head at CAPLE, and she was the national manager for the Cambridge English Portugal project in 2014. She has also co-authored English course books for all levels of learning. She holds a MA in Assessment from the University of Lisbon and has been developing her PhD project at the Institute of Education, University of Lisbon.

Ana Espírito Santo was a teacher of Portuguese as a Foreign Language for several years. She is a young researcher on L2 acquisition at the Centre of Linguistics of the University of Lisbon and a senior officer at CAPLE. She obtained her PhD from the University of Lisbon in General Linguistics (acquisition of relative clauses in PFL by Chinese learners) and her current research is focused on test item writing.

Are Humans Redundant? Automated Scoring in Learning and Assessment [D]

Dr. Arum Perwitasari

Artificial intelligence provides significant values in automated scoring in learning and assessment. Automated scoring offers performance-specific feedback, which is not feasible under operational human scoring and help score the responses efficiently and reliably, especially for test programs with large test taker volumes. What are the roles of humans in automated scoring? Are humans redundant, thus they should be replaced by machines? The aim of this session is to present the validity of automated scoring in assessing speaking and writing tasks of the TOEFL iBT® test as well as in selected test preparation tools. The session offers some background information about research and development of automated scoring systems at Educational Testing Service (ETS). It provides an overview of the automated scoring capabilities, including e-rater® automated scoring and feedback engine used for assessing writing and SpeechRater® service for assessing speaking and explains the design of the two to ensure the scoring quality. While automated scoring models have advantages, the limitations of automated scoring and assessment will also be discussed. This session will also highlight a hybrid approach in scoring constructed responses by combining human raters and automated scoring tools to maintain the validity, reliability, and fairness of the test scores. It sums that even though automated scoring in on all our future in learning and assessment, humans are still needed to attend to a wider variety of features, such as the quality of ideas and content as well as form in scoring responses.

Dr. Arum Perwitasari is an experienced educator, ELT enthusiast, academic editor and researcher. Her research interests include psycholinguistics, second language acquisition and bilingualism. She is an Institutional Relations and a content lead on language assessments, ELT and higher education research at ETS Global Amsterdam. Arum holds a Doctoral degree in Linguistics from Leiden University Centre for Linguistics (LUCL), the Netherlands.

What is the future of plurilingual language assessment for ‘monolingual’ testing organisations? [DI]

Graham Seed

In the era of the ‘multilingual turn’ in language education (May, 2014), academic critics have been quick to point out the lack, or slow speed, of recognition or adoption of plurilingual, code-switching and/or translanguaging practices within language assessment (e.g. Shohamy, 2011). Others have pointed out the difficulties in defining the construct of such an assessment as well as the challenge of educational policies in promoting these (Saville & Seed, 2021). One particular source of vexation is how language tests, so historically entrenched in promoting proficiency in languages along monolingual lines, can ever make a truly plurilingual turn. In Europe especially, it has only been in the last few years that practitioners have been able to identify what a ‘plurilingual assessment’ might look like in practice (such as De Angelis, 2021; Seed & Holland, 2020), often with the help of the Common European Framework of Reference for Languages (CEFR) Companion Volume’s attempts to provide reference level descriptors for plurilingual competence and mediation between named languages (Council of Europe, 2020).

In order to address the question of whether and how test-takers already utilise their plurilingual repertoires, this paper goes on to report on how examples of code-switching are found even within what are supposed to be monolingual written test responses, in a study of more than five hundred test responses taken from the automated “Write & Improve” learning-oriented assessment platform (Cambridge University Press & Assessment, 2022).

We evaluate what that might mean in practice before finally taking a more futuristic look at how the use of digital technology and AI, including automated assessment and feedback, provide the best chances of creating a truly personalised plurilingual assessment offer. In doing so, we review the nascent work already started (Nguyen, Yuan & Seed, 2022) and consider what the next steps might be in order to achieve more long-term aims.

Additional Authors: *Li Nguyen, Institute for Automated Language Teaching and Assessment (ALTA) and Zheng Yuan, King’s College London.*

Graham Seed is a Senior Research Manager at Cambridge University Press and Assessment, and the Secretariat Manager of ALTE. He holds a Master’s degree in Language Testing from

Lancaster University. His current research interests include language assessment in plurilingual contexts and the operationalisation of the CEFR in language assessment.

Aligning a multimodal integrated speaking assessment task to the Common European Framework of Reference [F]

Kim Anne Barchi and Mariana Jo Bisset

The present study is a pilot project aimed at aligning multimodal integrated speaking performances to the Common European Framework of Reference (CEFR Companion Volume, 2020) descriptors, increasingly accepted as a key point of reference for language test developers. Measures of discourse and linguistic features were extracted from the transcriptions of test takers’ performance and analysed.

University of Padova students attending the faculty of education take a mandatory computer-based speaking test which comprises 3 tasks. In particular, the present study focuses on the second task, which involves multimodal input: a video and short article stimuli to be incorporated into the speaking performance. Transcriptions of 85 recordings were created and analysed. Measures focused on idea units both from source input and test takers’ own; speech organization; discourse markers; indicators of grammatical accuracy and lexical variety. The data extracted was then compared to the selected B2 level descriptors of the CEFR in accordance with the test constructs.

The initial findings emphasise the multidimensionality of speech and the need to include a variety of measures in the development of rater scales for L2 speaking assessment. Moreover, analysis supports claims of test validity by providing evidence that the language elicited reflects the CEFR descriptors selected. Furthermore, considerations on pedagogical implications and suggestions for additional research are discussed.

Kim Anne Barchi is a Language Teaching Expert at the Language Centre of the University of Padua, Italy. She is part of the testing unit of the Language Centre of the University of Padua, and on the testing commission of the Language Centre of the University of Bolzano, Italy. She is attending the second year of a PhD at the University of Leicester, U.K., researching multimodal integrated assessment. Her research interests focus on oral testing and the wash back effects.

Mariana Jo Bisset is English Language Expert at the Language Centre of the University of Padua, Italy. She coordinates the testing unit of the Language Centre, developing in-house language tests and supervising the entire testing validation cycle and testing administration.

An Educational Action Plan to Improve Assistance to Autistic Spectrum Disorder (ASD) test-takers in written large-scale exams [DI]

Anarcisa de Freitas Nascimento

In Brazil, the National High School Exam (Exame Nacional do Ensino Médio – ENEM) is the responsibility of the Anísio Teixeira Educational Research Institute (Inep), which is part of the Ministry of Education. The growing presence of test-takers with Autistic Spectrum Disorder (ASD) and the need of specialized assistance in the writing component of the Exam inspired the present study. The aim of this presentation is to show analyses of accommodation procedures offered to the demands for specialized attendance in writing for students with ASD in Canada, United States and United Kingdom, and highlight the linguistic characteristics of such candidates based on the analyses of texts produced by Enem 2017 ASD participants. The two types of analyses – the experiences abroad and the textual analysis – resulted in an Educational Action Plan for Inep’s management teams so as to improve the training plans for readers and transcribers of the exam for students with autism, the improvement of the logistical support of the exam, with the inclusion of the professional Oral Language Modifier and differentiated ambience resources, the application of a written test on a computer and the inclusion of linguistic resources such as pronominal inversion, phonological writing, echolalia and textual text clipping, typical characteristics of students with Autistic Spectrum Disorder (ASD), empirically verified in the sample and essays, in the differential correction blueprint for ASD.

Additional Author: *Gladys Quevedo-Camargo, University of Brasília.*

Anarcisa de Freitas Nascimento graduated in Letters_Portuguese and English languages and specialized in Applied Linguistics. She holds an MA in Educational Management and Assessment in Public Education, with emphasis on exams and special needs. She works at National Institute of Assessment and Statistics in Brazil and gives pedagogical support in languages to Brazilian large-scale exams. She also teaches English in a public school.

Reframing language learning: Introducing the Sustainability Framework for ELT [DI]

Helen Allen

As a focus on sustainability becomes accepted at governmental and institutional levels and more visible within local and global communities, educators are striving to bring sustainability into their curriculum. Young people need, and want, to play an active part in shaping their world, and bringing sustainability education into the language classroom can help learners engage with and benefit from their language learning both in the now and for the future. In this session, we'll be presenting the Sustainability Framework for ELT (English Language Teaching), developed by Cambridge University Press & Assessment to guide teachers, publishers and other supporting stakeholders in developing suitable approaches to the development of sustainability skills within language education. Based on extensive research, and with the Cambridge Life Competencies Framework as a key point of reference, the Sustainability Framework aims to develop informed, compassionate and innovative agents for positive environmental, social and economic transformation. We'll explore dimensions of Knowledge, Values, Innovation and Transformation and competencies that range from 'critically evaluating sustainability claims' to 'generating hope in self and others'.

Our research has highlighted both challenges and opportunities, many of which hinge around the development and assessment of skills and the need for contextualisation: making space for both local and global cultural perspectives and the affordances of plurilingual practices. We'll be sharing insights into teacher and learner responses to the framework and related resources, and looking at what this means for how we approach the teaching, learning and assessment of sustainability skills in ELT within diverse contexts.

Additional Authors: *Jasmin Silver, Cambridge University Press & Assessment, Delia Kidd, Cambridge University Press & Assessment, and Jade Blue, Freelance Consultant.*

Helen Allen leads on insight application within the Product, Insight & Ideation team at Cambridge University Press & Assessment. She works in a cross-functional team in the investigation, development and integration of a range of language learning products – to nurture meaningful engagement and improved learning outcomes.

Implementing CAF measures and the scales of Functional Adequacy in the L2 Italian certificate exam Cert.it: Implications for the assessment of B1 candidates applying for citizenship [F]

Diego Cortés Velásquez and Elena Nuzzo

Since 2018, it is required by the law that applicants for Italian citizenship provide evidence of their linguistic proficiency at the B1 level or higher. Consequently, in Italy like in other European countries, several institutions have adjusted their B1 certificates to the needs of this specific population of learners (Deaiana, 2021).

The L2 Italian certificate exam Cert.it B1-citizenship consists of closed-ended questions for receptive skills and open-ended questions for oral and written production. Currently, the open-ended answers are scored using scales predominantly based on accuracy, whereas less attention is devoted to other aspects of language use, such as task fulfillment and pragmatic appropriateness. Still, there is growing agreement that these aspects of language proficiency need to be considered when assessing L2 users, in line with the task-based approach recommended by the European institutions (cf. Council of Europe, 2001; ALTE, 2016). This might be particularly relevant for the assessment of learners who developed the target language in migration contexts.

Therefore, Cert.it is planning to revise and improve its assessment tools by applying findings from research in the area of Instructed Second Language Acquisition. In this study, we explore the implementation of the CAFFA (Complexity, Accuracy, Fluency, and Functional Adequacy) framework (Kuiken and Vedder, 2022) on written productions. Two groups of raters are asked to rate a sample of written tasks performed by Cert.it B1-citizenship candidates. One group relies on the scales currently used by Cert.it whereas the other refers to a selection of the CAFFA measures, adapted for this study. The presentation will focus on the comparison between the results of the two rating procedures.

References: ALTE (2016). I test linguistici per l'accesso, l'integrazione e la cittadinanza: Linee guida per i responsabili politici, Consiglio d'Europa, Bruxelles. Council of Europe (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment, Council of Europe Publishing, Strasbourg. Kuiken, F. & Vedder, I. (2022). The Assessment of Functional Adequacy in Language Performance: Special issue of the Journal on Task-Based Language Teaching and Learning, 2:1. Deiana, I. (2021) I test

linguistici per l'integrazione e la cittadinanza: alcune riflessioni sul caso italiano. Italiano LinguaDue, n. 1. 2021.

Diego Cortés Velásquez is an associate professor of second language instruction in the Department of Foreign Languages, Literatures, and Cultures at Roma Tre University. His current research includes issues related to multilingualism, Task-Based Language Teaching, and cross-cultural pragmatics. He is currently a member of the scientific board of Cert.it.

Elena Nuzzo is an associate professor of second language instruction in the Department of Foreign Languages, Literatures, and Cultures at Roma Tre University. Her work focuses on research, education, and teacher training in the field of applied linguistics, with a specific interest in Italian as a second language. Her main areas of research include the practical applications of Speech Act theory in the learning and teaching of second languages, cross-cultural pragmatics, and task-based language teaching. She is currently a member of the scientific board of Cert.it.

Automated Scoring of Spelling Mistakes in Short Answers [D]

Leska Schwarz and Ronja Laarmann-Quante

Advances in the field of language technology open up new perspectives for the development of automated scoring systems in language assessment. In this paper, we investigate to what extent test-takers' short answers on the listening tasks of the digital TestDaF (digitaler Test Deutsch als Fremdsprache, Test of German as a Foreign language) can be scored in an automated way.

In this corpus study, we analyse human decisions regarding the acceptability of very short misspelled responses of one or two words in length. Responses are rated by trained human raters according to scoring guidelines for each item. Misspellings of the target answer are marked as correct if the response remains comprehensible and shows sufficient understanding of the input listening text. Despite the guidelines, raters often find it difficult to judge whether a response containing one or more orthographical errors is to be marked as correct or incorrect.

In the study, we investigated how technological models can be developed to explain why an answer is classified as correct or incorrect. We explored possibilities to operationalise these decisions using features that could be applied by an automatic scoring system, using measures like character edit distance and

phoneme similarity between the target answer and answers given by test takers. The goal was to model acceptability decisions in such a way that decisions can be explained and automated by language technology.

The results show that the automated scorings reach performances that are close to human inter-rater agreement, but cannot yet explain all decisions. We discuss the remaining inconsistencies and perspectives for future research in this field.

Additional Authors: *Andrea Horbach and Torsten Zesch.*

Leska Schwarz obtained her Bachelor's degree in Instructional Design and German Linguistics and her Master's degree in Educational Science from the University of Freiburg im Breisgau, Germany. She's currently a test validation officer at the TestDaF Institute in Bochum, Germany. Her research interests include the assessment of the receptive skills with specific focus on the testing of listening comprehension, and the possibilities of automated or technology-assisted scoring of short-answers.

Dr. Ronja Laarmann-Quante studied Linguistics and obtained her PhD in Computational Linguistics from Ruhr University Bochum, Germany. She is currently working as a postdoctoral researcher at CATALPA – Center of Advanced Technology for Assisted Learning and Predictive Analytics at the FernUniversität in Hagen, Germany. Her research focuses on the use of language technology in the educational domain in order to support language learning.

Language assessment across a hundred languages - how to ensure test quality and valid results? [DI]

Mari Nielsen Vaage and Marius Lian

Interpreting is often a prerequisite for overcoming language barriers in the public sector, and to ensure legal safeguard, equality and democratic rights and duties. In Norway, there is a registered need for interpreters in more than a hundred languages. Oslo Metropolitan University (OsloMet) is the leading academic and professional environment when it comes to interpreter education and language testing in Norway, offering a comprehensive set of education and qualification schemes for interpreters at all levels of The Norwegian National Registry of Interpreters. In addition to two BA programs in sign language interpreting and in public sector interpreting, offered in a multitude of spoken languages, OsloMet

offers lower-level testing of interpreters through The Bilingual Proficiency Test (Tospråktesten), currently developed in nearly 80 languages, and higher-level testing of interpreters through the Norwegian interpreter certification exam offered in more than 40 languages. No other university in the world educates and tests interpreters in as many languages as OsloMet.

Tospråktesten is an oral test functioning as an entry test to both OsloMet's interpreter course as well as to the bachelor program. Tospråktesten aims to test that the candidates have the minimum necessary language skills to start a qualification process as an interpreter. The Certification Exam, on the other hand, marks the highest level of professionalism and is aimed at practicing interpreters. Both tests are considered high stakes tests and gate keepers for the interpreter profession. Working across so many languages, many of which are not fully standardized, how do we ensure validity and reliability? As a new associate member of ALTE, OsloMet would like to present the work we do within language assessment with a specific focus on the two tests and how we recruit and train our examiners.

Additional Author: *Mona Myran*

Mari Nielsen Vaage is a senior advisor at Oslo Metropolitan University and works with language testing and the qualification of interpreters for Norway's public sector. She mainly works with the administration and development of Tospråktesten – a bilingual proficiency test for potential interpreters – and TAO – an introductory course to the interpreter profession. Mari holds a Master of philosophy in Organization, Leadership and Work and a BA in Culture and Communication. She also has experience from research within the field of migration and diversity.

Marius Hemstad Lian is a senior administrative officer at Oslo Metropolitan University and works with language testing and the qualification of interpreters for Norway's public sector. He is mainly involved with administering the National Interpreter Certification Exam (NICE). Marius holds an MPhil from the Centre of Development and the Environment (University of Oslo) and a BA in Asia and Middle East Studies (University of Oslo).

Alignement d'un test d'évaluation de français langue étrangère sur des cadres de référence distincts [F]

Dominique Casanova et Emilie Grosset

Le Cadre européen commun de référence pour les langues (CECRL), dans sa version française, et les Niveaux de compétence linguistique canadiens (NCLC) sont deux schémas descriptifs de la compétence langagière d'usagers de la langue française pour qui le français ne constitue pas la langue première. Ces cadres ont tous deux connus des évolutions en 2018 (avec, pour le CECR, la parution du Volume complémentaire). L'échelle québécoise des niveaux de compétence en français des personnes migrantes adultes (EQNCF) est pour sa part en cours de bonification.

Si les changements apportés s'inscrivent dans la continuité des cadres existants, les versions actualisées méritent d'être prises en considération en cas d'évolution du format de tests dont les résultats sont interprétés en référence à leurs échelles de niveaux. C'est le cas du Test d'évaluation de français – TEF, qui est notamment utilisé dans le cadre des démarches d'obtention de la carte de résident ou de la citoyenneté au Canada, en France et au Québec. Le Français des affaires, qui fait évoluer le format des épreuves de compréhension écrite et de compréhension orale du TEF, a donc entrepris un nouvel alignement de son test sur le CECR et les NCLC, s'appuyant pour cela sur la démarche proposée dans le manuel Aligning Language Education with the CEFR : a Handbook (2022).

Ces travaux d'alignement sont l'occasion de poser un regard critique sur la démarche de définition des points de césures en recourant à des panélistes statuant sur le degré de difficulté des items et sur des propositions faites dans Kaftandjieva (2010). Les résultats sont également à mettre en regard des travaux d'alignement direct des deux cadres entrepris par North & Piccardo (2018).

Dominique Casanova is Head of scientific development at Le français des affaires. He started working in the field of language testing in 2005, when coordinating the development of the electronic version of the Test d'évaluation de Français. In 2013-15 he managed the renovation of the IT system of Le français des affaires. He currently supervises psychometrics and projects involving machine learning and NLP. He is one of the chairs of the ALTE Special Interest Group on Technology in Language Assessment.

Emilie Grosset is an Education Manager at Le français des affaires/ French for Business, CCI Paris Île-de-France Education. She is in

charge of the conception of the Test d'Evaluation de Français, she contributes to the conception of Diplomas in French for Professions and liaises with examination centres on training and assessment matters. She is involved in the creation of education resources in partnership with RFI Savoirs, and participates in training sessions for French teachers in Didactics of French for specific purposes.

Adapting online language activities and assessment for functionally diverse learners [DI]

Jackie Robbins

Online learning institutions offer a great potential for all learners, but particularly for learners who may struggle in traditional learning environments. At the Open University of Catalonia, a growing number of learners are undertaking courses who have differing degrees of visual, hearing, psychological, physical and cognitive impairment and as almost all degree programmes at our university are required to demonstrate a B2 level, the Centre for Modern Languages has often been the first area when learners have needed learning activity and assessment adaptations. In this presentation, we will share the results of an ongoing project whose aim it is to continue to improve the adaptations for learners with diverse needs. We will identify the main reasons why our learners choose to study online and consider the key concerns involved when it comes to adapting online language courses for learners with diverse and often highly individual needs. We will also discuss the implications that functional diversity may have on assessment.

Jackie Robbins holds a PhD from the Universitat Autònoma de Barcelona and is a lecturer of English at the Centre for Modern Languages at the Universitat Oberta de Catalunya. Her main lines of research are in the field of engagement of language students in digital contexts. Her other areas of research interests include pedagogical innovation in online language learning, the role of teachers in the online language learning process and adapting online language learning for learners with functional diversity.

The use of Global Englishes in Listening assessment: A review of the evidence and future directions of research [DI]

Amy Devine

Test takers live in multicultural, multilingual environments and often aspire to use their English qualifications to move to environments that will also be multicultural and multilingual. Listening tasks, however, do not always reflect this reality, often only including very few English varieties for their input speech. This has prompted an interest in the development of more 'authentic' Listening tasks, that more accurately reflect the academic, work and social environments where English is used (Ockey & Wagner, 2018). This paper will discuss a synthesis of available research on the use of Global Englishes, i.e. Englishes spoken by native speakers (EL1) as well as by speakers who use English as a second, foreign or additional language (EAL) (Rose & Galloway, 2019), as input for Listening tasks in standardised English language testing. Relevant research was identified through the querying of academic databases and the forward- and backwards-tracing of key publications. Two researchers independently reviewed and coded identified research and subsequently extracted data from entries that were selected for inclusion in the review. The findings of the selected studies were analysed in order to understand whether the inclusion of varieties of English that are not typically represented in standardised English language assessment, including both the accents of speakers of World Englishes (i.e. EL1 varieties that are less included) and EAL accents, affects the reliability, validity and fairness of the assessment, as well as test taker performance and their perceptions of the task. The results of this review provide an overview of the empirical evidence regarding the use of Global Englishes in Listening task recordings and highlight areas of research that need to be further explored in order to understand how we can design more inclusive Listening tasks.

Additional Author: Carla Pastorino, Cambridge University Press & 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 digital products and examinations for young learners. Amy's 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, Amy worked on several funded research projects in education and psychology at the University of Cambridge. Amy completed a PhD in Psychology at the University of Cambridge in 2018.

La formación y actualización en línea de los calificadores del DELE [D]

Eva María García García

En esta ponencia se presentan las posibilidades de formación y actualización en línea de calificadores de sistemas certificativos, y se analiza la combinación de actividades autónomas y en grupo en función del momento y del objetivo que se persigan. Para ello, se muestra el ejemplo de los cursos de formación de calificadores del DELE. El examen DELE es un examen de dominio del idioma español que incluye dos pruebas de calificación subjetiva: la prueba de Expresión e interacción escritas y la de Expresión e interacción orales. Debido a la distribución de los calificadores en distintos continentes, la formación se realiza en línea y de forma asincrónica a través de la plataforma Moodle.

Los cursos en línea de calificadores del DELE siguen las fases de la formación recomendadas en el Manual para vincular exámenes al MCER: familiarización con los niveles y las escalas, ilustración de las bandas de las escalas y calificación de muestras reales. El primer curso de un calificador DELE está tutorizado e incluye actividades interactivas de respuesta automática, actividades de ilustración de respuesta abierta corregida por el tutor y foros de debate para discutir en grupo y de forma tutorizada aquellas calificaciones en las que haya menos consenso.

Por su parte, los cursos de actualización que los calificadores deben realizar de forma periódica son de realización autónoma y tanto las actividades de ilustración como la calificación de muestras tienen retroalimentación automática.

En la presentación se describirá el diseño, la elaboración y el funcionamiento de los cursos, se compartirán resultados de encuestas a los calificadores sobre la formación y se extraerán conclusiones sobre su eficacia.

Eva María García García es licenciada en Filología Hispánica por la Universidad Complutense de Madrid. Trabaja en el Instituto Cervantes, donde comenzó en la Oficina del Español en la Sociedad de la Información, en proyectos de difusión de tecnologías lingüísticas en español. Después se trasladó a la

Unidad de Certificación Lingüística donde coordinó proyectos para desarrollar herramientas tecnológicas aplicadas a la evaluación: Pruebas Adaptativas de clasificación y Certificado Internacional de Español. En la actualidad es miembro del equipo que realiza la gestión académica de los exámenes DELE.

Calibration of a “Golden corpus” to assess the effectiveness and validity of an algorithmic solution (Artificial intelligence) for assistance of writing assessment : findings, limits, opportunities [D]

Vincent Folny

TCF (Test de connaissance du français) is a test developed by France éducation internationale (FEI) a french agency (Ministry of Education). This test has been developed to deal with reliability and validity by design. To be selected and hired raters are tested. Annually FEI produces rater reliability indices based on all live data and a commission selects raters for the next round of 12 months. A double rating is applied systematically. Since 2014 FEI used a computerized solution for writing. A massive corpus of oral and written productions doubled rated and linked to reliability indices is available. In 2020, FEI received public funds from FTAP (Fonds de Transformation de l'Action Public) to modernise the assessment of writing. It has been decided to work on a partial automation of computerised ratings and on the assistance of the rater. As the training corpus used for Artificial Intelligence automation is large and characterized, it offers an opportunity for a partial automation solution.

The evaluation of the algorithmic solution efficiency used to rate the candidates is more challenging. The use of double rated productions (live test data) is not so relevant. FEI decided to create a “Golden Corpus” for validation purposes. 1083 performances have been carefully selected. 56 TCF raters have been involved. Each production has been rated at least by 9 raters and 10 % of the performances rated by all raters. A multifaceted Rasch analysis has been run and its results taken in account for the final selection of the performances. During this presentation, the result of the corpus calibration will be analysed as well as the reliability of raters and performances. Given that this project is innovative (new and risky) and TCF used for its robustness, we will explain how

we intend resolve this apparent contradiction based on findings, argumentation, and responsibility.

Vincent Folny is in charge of innovation applied to French exams in France éducation internationale (FEI). He develops new exams, and he is involved in new developments of DELF-DALF and TCF. He is currently working on the development of an automated scoring of essays for the Test de connaissance du Français (TCF). Vincent Folny is actively involved in the definition and implementation of ALTE's quality management. He chairs the ALTE Standing committee, conducts regulatory audits in Europe and train auditors.

Putting the CEFR at the centre of classroom assessment in a global teaching program [F]

Aidan Holland, Carolyn Westbrook and Johnathan Cruise

Since its publication in 2001, the CEFR has generated a strong body of good practice in alignment to standards for assessment. Courses in teaching and learning contexts, however, have not always matched this evidence-based approach to alignment. In 2022, a cohort of international collaborators published a Handbook on Aligning Language Education with the CEFR to offer a corrective to this paradigm. This presentation reports on a researcher-teacher collaboration to implement ideas in the Handbook to improve a large-sale classroom language program offered across multiple countries and regions for secondary school-age students. The programme is structured around sets of lessons targeting a CEFR level. Each set includes thematically-linked tasks and activities culminating in a project.

The assessment approach helps teachers evaluate students' progress in a meaningful way that is also demonstrably linked to the CEFR. A cross-disciplinary team of researchers and teachers first mapped the content of courses to the CEFR using activities in the lessons. This was then turned into a checklist approach for teachers to observe and evaluate student performance on the spoken output in the projects. These assessment tools have been piloted in a range of teaching centres across Europe and Asia in order to build a toolkit for a global roll-out of the approach. The project has followed an iterative process of collaborative evaluation and revision of the tools for assessment. An important focus of the researcher-teacher project team has been to achieve an appropriate balance between detail and feasibility for teachers working in a

busy language-school context with limited time for preparation and marking of assessments outside of class time. This paper will describe the process of ensuring a principled and documented approach to mapping CEFR descriptors to lessons and provide results of piloting this approach in a global teaching program.

Additional Author: Jamie Dunlea, British Council.

Aidan Holland has been working in English language teaching since 2004. He worked in Taiwan Italy and the UK before settling in Spain in 2009, where he began work for the British Council. He has worked in a range of contexts and has experience with both young learners, teenagers and adults. Since starting with the British Council, he has worked in a variety of roles at a local, national, regional and global level. He currently has an MA in TEFL/SL from the University of Birmingham. Has a particular interest in language assessment testing and is currently studying a second MA in Language Testing with Lancaster University.

Carolyn Westbrook is a Test Development Researcher at the British Council in the UK. She is a Senior Fellow of the Higher Education Academy in the UK and was formerly an Associate Professor in EFL. She has over 25 years' experience teaching and assessing General English, Business English, EAP and ESP, as well as working as a teacher trainer and materials writer. Carolyn's main focus of research is integrated skills assessment and she has taken part in a number of testing and assessment development projects. She has also worked on major projects aimed at developing language teachers' assessment literacy. After obtaining a BA(hons) in Literature and Linguistics from the University of Wales and a PGCE in General Primary Education from Bedford University in 1993,

Johnathan Cruise moved to China to work as a teacher trainer and examiner. In the last 12 years, Johnathan has worked in test development and examiner management and also received an MA in Applied Linguistics and TESOL in 2010 from the University of Leicester with a dissertation focus on speaking tests. He is currently working in the UK with the British Council's Assessment Research Group and will start on a PhD on image use in YL tests in 2023.

Developing standardized tests for Arabic for the final exams in secondary education: experiences in the Netherlands [DI]

Anneke de Graaf

In the Netherlands, secondary education ends with national examinations. One of the official languages that can be chosen on the different levels is Arabic. Since the nineties, Cito develops standardized tests for reading comprehension in Arabic. The exams for the five levels in secondary education consist of a reading comprehension test about authentic texts from all over the Arabic world. Students need to apply different reading strategies to be able to answer questions which have different formats, depending on the level of the exam.

This presentation shows the challenges that the test development groups still have to tackle when developing these tests for Arabic. Arabic has a long history in Dutch education at an academic level, but how should requirements and standards for language assessment be transferred to the context of Modern Standard Arabic in the 21 century European society? Can we find for example texts in the media to be used for the lower levels while preserving quality of language? And, concerning the (CEFR-) level, should we aim at Moroccan children living in the Netherlands, at students without any background in this language or at refugee students coming recently from the Middle East? Another important issue is: can we align the exam to other language exams in the Netherlands and also to CEFR?

For the various aspects of the test development cycle, tools were developed, then reviewed and adapted for the goal of developing standardized exams for Arabic. Checklists, capacity building, protocols and CEFR standard setting procedures helped in developing the quality of these standardized exams for Arabic to give this language a better position in the Dutch educational context.

Anneke de Graaf works as a senior consultant in the domain of language assessment at Cito on a national and on an international level. She is also involved in the development of national language exams in the Netherlands.

Assessment in the Early Years: Mapping Concepts and Practices in four Brazilian states [DI]

Juliana Reichert Assunção Tonelli

Brazilian primary and secondary education is ruled by the National Common Curriculum Guidelines (BRASIL, 2018). According to this document, English is a compulsory discipline from the 6th to the 9th grades (elementary school) as well as in the three years of secondary school. However, there has been a considerable growth in the offer of English in the curriculum in the early years of elementary school (1st to 5th grades) in many Brazilian municipalities. With no official guidelines for this educational phase, each municipality organizes its curriculum based on their own concepts and resources. In 2021 a one-year study was carried out by a group of teacher educators and researchers from five Brazilian universities in four Brazilian States. This study, supported by the British Council, aimed at presenting alternatives for the teaching of English in the early years as well as to contribute for the accessibility and democratization of the teaching of English as an additional language to children in public schools, thus enabling linguistic education at this level. Three specific objectives were established: 1) to map the municipalities in the Brazilian States that offer English in the early years of elementary school; 2) to produce a document support the development of public policies for that teaching context as far as English is concerned; 3) design a teaching unit to be sent to children in vulnerable situations and/or with difficulties of access to materials to learn English. Based on the first specific objective, this presentation will focus on the data collected by means of an online questionnaire sent to the municipalities' secretariats of education in the four States, particularly with respect to assessment types, criteria and methodology used. The findings are concerning as, out of the 1.369 municipalities investigated, only 46 mention assessment types or methodology in their curricula.

Additional Author: *Gladys Quevedo-Camargo, University of Brasília.*

Juliana Reichert Assunção Tonelli holds a PhD in Language studies from Londrina State University). She is an associate professor at the Department of Foreign Languages at Londrina State University, where she works with English language teacher education at undergraduate and graduate levels. Since 2000 she has devoted her time to researching the teaching of additional languages to young learners from pre-school to 5th grade.

The Comparability of Computer-Based and Paper-Based Writing Tests: A Case Study [D]

Balázs Csizmadia

The aim of this paper is to report on research into the comparability of computer-based (CB) and paper-based (PB) Writing tests. It analyses the two types of Writing tasks in English used by ELTE Origó Language Centre at B2 level. The PB and CB Writing tests were identical except for the mode in which they were administered. The study examines possible differences in candidate performance and test-taking experience resulting from the mode of delivery, and compares the validity and reliability of the two types of test. Candidates were divided into two groups, with half of them asked to do Writing Task 1 on paper and Task 2 on computer, and the other half vice versa. Although not based on a great number of samples, this study aims to offer some insights into the similarities and differences between computer-based and paper-based Writing tests that are relevant to other types of language examinations as well.

Balázs Csizmadia holds an MA in German and English Studies from Eötvös Loránd University, Budapest (ELTE), and a PhD in English from Royal Holloway, University of London. Currently working for ELTE Origó Language Centre in Budapest, he is relatively new to the field of language testing and assessment. However, he has attended several ALTE conferences and one EALTA conference, and he has also been working as a language teacher for over twenty years.

Balancing the need for native and non-native speakers in ELF listening tasks: to what extent do accents affect comprehension? [DI]

Anna Maria De Bartolo and Jean Marguerite Jimenez

In a rapidly changing world in which English has achieved a status of an international language and is spoken as a lingua franca (ELF) by growing numbers of non-native speakers in academic, cultural, educational and professional domains, the way it is assessed as an L2 has become a topic of heated debate, bringing into question

whether native speaker norms should still be considered the standard in language assessment, especially when more attention is being placed on the intercultural and plurilingual aspects of language learning (CEFR, 2020). Although calls for the development of tests which reflect the communicative and contingent variability nature of ELF have been largely unheeded by international language examination certifiers (Jenkins & Leung, 2016), a careful rethinking of a test construct which takes this variability into account warrants further investigation.

Turning specifically to the assessment of listening, this would imply a move away from standard English accents towards an inclusive approach which focuses on the use of both native and non-native speaker accents (see Abeywickrama, 2013; Harding, 2008, 2011, 2012; Kang et al., 2018; Major et al., 2002; Newbold, 2017). This may be particularly important in tertiary education, where students are likely to encounter a variety of accents throughout their course of study. In light of this, the present study investigates the extent to which B1-level Italian L1 speakers' performance on listening tasks might be affected by the speaker's accent. Specifically, undergraduate students from different degree courses will be assessed on their comprehension of five extracts of a lecture delivered by fluent speakers of English with different L1 accents (American English, Arabic, British English, Hungarian, Italian). The test tasks are similar regarding topic, speech rate and syntactic complexity and will be delivered to the participants in a random order. Preliminary results will be presented and pedagogical considerations suggested.

Additional Author: *Ian Michael Robinson, University of Calabria.*

Anna Maria De Bartolo is Associate Professor of English Language and Translation at the University of Calabria, Italy, where she teaches English for Administration Science and English for Pedagogical Studies. She holds an MA in Modern Literature and Culture, York University, UK and an MA in Language Teaching: Applied Linguistics, Lancaster University, UK. Since 2005 her research interests have mainly focused on ELF studies and the issues surrounding standard/non-standard varieties of English. She has published in the area of multilingualism and multiculturalism and is interested in exploring the implications of ELF for English language policy, pedagogy, and practice as well as ELF in relation to Intercultural Communication.

Jean Marguerite Jimenez is Associate Professor of English Language and Translation at the University of Calabria, Italy. She is part of the Advisory Board of the University Language Centre, where she is co-coordinator of the Test Development Team responsible for the design and development of EFL tests for university students. She holds a PhD in Applied Linguistics from Lancaster University, UK. Her research interests include Second Language Acquisition,

Corrective Feedback in CALL, Language Testing, and the use of Corpus Linguistics in the second language classroom. She has presented papers at national and international conferences in Europe and North America.

El diseño interactivo del examen del SIELE [D]

Aurora Duque de la Torre

Al diseñar un examen certificativo de lenguas por ordenador, hay que tener en cuenta las consecuencias que las decisiones tomadas tienen sobre la validez, la fiabilidad, la interactividad, la autenticidad, y la experiencia del usuario del examen, así como las diferencias con respecto a exámenes en papel, si se trata de un examen con dos versiones.

En esta comunicación se presentan las diferentes decisiones de diseño tomadas en la creación del examen certificativo de español SIELE, un examen multinivel (A1-C1) por ordenador compuesto de cuatro pruebas: Comprensión de lectura, Comprensión auditiva, Expresión e interacción escritas y Expresión e interacción orales.

Para la administración del examen SIELE se diseñaron una plataforma y unas tareas de carácter interactivo que debían adaptarse a la idiosincrasia de cada prueba y garantizar al mismo tiempo la fiabilidad de la administración del examen y la facilidad de uso.

En la ponencia, por un lado, se analizarán las decisiones tomadas con respecto al diseño de la interfaz y la interacción con el sistema en cada una de las pruebas, de acuerdo a la idiosincrasia de las mismas: las dinámicas de las tareas, la gestión automática del tiempo, la navegación entre tareas, la disposición del espacio en pantalla, la incorporación de los caracteres específicos del español para otro tipo de teclados, o la activación automática de los audios y la grabadora.

Por otro lado, se explicará el funcionamiento del sistema de selección de tareas para cada modelo individual de examen, que tiene un motor aleatorio con restricciones de combinación en función del tema y la variedad del español empleada en cada tarea.

Finalmente, se extraerán conclusiones sobre las decisiones tomadas y sobre aquellos aspectos que, al tratarse de un examen por ordenador automatizado, supongan una diferencia en las propiedades del examen con respecto a pruebas administradas por personas.

Additional Authors: *Juan Miguel Prieto, Sandra Cruz y Goergina Lacanna.*

Aurora Duque de la Torre. Licenciada en Filología Hispánica por la Universidad Complutense de Madrid y máster de ELE por la Universidad de Alcalá. Ha trabajado como profesora de ELE en la Universidad de Alcalá y en la Escuela Complutense de Verano. Ha participado en la elaboración de cursos de español por Internet: Aula Virtual de Español; ¡Hola, amigos!; Hola, ¿qué tal?, etc. Trabaja en el Departamento de Certificación Lingüística en el proceso de diseño, creación y estandarización de los exámenes DELE y, actualmente, del examen SIELE. Es formadora de examinadores de los cursos de acreditación de examinadores DELE. Es coautora del manual Agencia ELE 2.

Considering inter-test relationships in high-stakes admissions testing: The case of English [D]

Ramsey Cardwell and Steven Nydick

High-stakes language assessments are moving to fully-digital platforms, an early example being the Duolingo English Test (DET). Test developers must demonstrate that online assessments are equivalent to in-person alternatives, as many institutions accept multiple assessments to demonstrate English proficiency. Evidence of construct similarity typically includes inter-test score correlations and concordance analyses. Total-score correlations >0.7 have been reported for TOEFL–IELTS (ETS, 2010), DET–IELTS, and DET–TOEFL (Cardwell et al., 2022). This study investigates observed inter-test correlations over time and updated inter-test concordances for English assessments.

DET examinees voluntarily report external assessment scores for research purposes. As an online on-demand assessment, a continuous data stream allows for real-time monitoring of inter-test relationships (Liao et al., 2021). Yet the demographics of examinees who report external scores, as well as the correlation between the DET and other assessments, appear to vary between January 2020 and May 2022. Multiple factors can impact observed correlations, including sample size, test reliability, and ability distributions. We applied two methods—multiple imputation (Buuren, 2018) and data weighting—to account for population changes in demographics and ability. Both methods eliminated observed fluctuations in correlations, implying that the observed change does not reflect anything substantive about the relationship among the tests.

Although correlations appear stable, score scales might still drift over time. We thus conducted analyses to update the DET–IELTS and DET–TOEFL concordance tables, which facilitate consistent and fair admissions decisions. We obtained official IELTS/TOEFL score reports from DET examinees (TOEFL n=328; IELTS n=1,643) and supplemented these data with the aforementioned self-report data (adjusted for small reporting error). We applied kernel equating and conducted sensitivity analyses to confirm the results' stability. We will discuss the implications of the concordance results for test developers and test users, the primary implication being that inter-test relationships should be monitored over time.

Additional Author: *J.R. Lockwood, Duolingo.*

Ramsey Cardwell received his PhD in educational measurement from the University of North Carolina at Greensboro, where he focused on validity theory and language testing. He completed three internships in psychometrics and educational measurement during his graduate studies and has been with Duolingo since completing his PhD. Ramsey is currently on the validation team of the Duolingo English Test.

Steven Nydick received his PhD in quantitative psychology from the University of Minnesota, where he was a Dissertation Fellow and studied item selection algorithms and stopping rules for adaptive mastery tests. He has almost 10 years of professional experience finding practical and defensible solutions for building high stakes exams across a wide range of purposes. Steven currently applies his knowledge to both the validation and test-scoring teams for the Duolingo English Test.

Using the CEFR descriptors as Performance Level Descriptors (PLDs): Opportunities, challenges, and insights [F]

Paraskevi (Voula) Kanistra

Aligning a test instrument to a global proficiency framework customarily entails panellists using performance level descriptors (PLDs) or achievement level descriptors (ALDs) when recommending cut scores. Such PLDs or ALDs explicitly state and quantify the knowledge, skills, and abilities (KSAs) test takers need to pass a test. In the case of a CEFR alignment study, the CEFR descriptors

can be used in the place of PLDs. However, the CEFR descriptors are qualitative rather than quantitative as they describe “what test takers can do” at each CEFR level, but do not mandate “how much test takers should do”.

As part of my PhD study, I investigated panellists' perceptions of using the CEFR descriptors as PLDs when recommending a CEFR cut score on a mono-level English as a foreign language examination. Within a week of the virtual workshop, all nine panellists participated in one of three focus group interviews. The interviews were transcribed, and the data were coded using inductive and deductive approaches.

In this presentation, I will focus on one of the themes that emerged from the focus group data analyses: Using the CEFR instead of Performance Level Descriptors (PLDs). In all three focus group interviews, panellists reported on the associated opportunities offered and challenges faced with using the CEFR descriptors. Panellists provided insights into the techniques employed to “tame the CEFR”. The panellists' perceptions were analysed from multiple perspectives reflecting the composition of the panel (novice vs. experienced; internal vs. external). To triangulate their perceptions, the panel discussion between Round 1 and Round 2 of the virtual workshop was transcribed and analysed. The analysis revealed that panellists used the CEFR descriptors substantially when rationalising their judgments.

The research findings presented will be valuable to standard setting practitioners wanting to use the CEFR descriptors as PLDs in CEFR alignment studies.

Paraskevi (Voula) Kanistra is the Head of Technical Content at Trinity College London. She leads a team of assessment specialists who manage the creation and quality assurance processes associated with Trinity's exams. She has a background in English Language Teaching and over two decades of experience in assessment design and validation. She presents regularly at professional conferences and is particularly interested in issues relating to standard setting methodology. She is currently completing a PhD at the University of Bremen.

Effects of examinee L1, assessment criteria and L1 recognition in Finnish National Certificates of Language Proficiency (NCLP) [DI]

Tuija Hirvelä and Reeta Neittaanmäki

When assessing speaking performances, even qualified, trained and experienced raters can be affected by different kind of attitudes and stereotypes towards various first language (L1) speaker groups. Consequently, recognition or assumption of the examinees' L1 might impact negatively or positively rating behaviour. In order to guarantee fair assessment regardless of speakers' L1 in a high-stakes proficiency test, National Certificates of Language Proficiency (NCLP), interactions between L1, (mis)recognition of the L1 and rating of proficiency in seven criteria have been studied in project 'Broken Finnish': Accent perceptions in societal gatekeeping (Academy of Finland 2018–2022). In this presentation we will introduce our modeling of the three-fold interaction between L1, seven criteria and the L1 recognition using Cumulative Link Ordinal Regression Models in R.

The data of the research consist of altogether 49 speech samples of 10 Arabic, 10 Estonian, 9 Finland Swedish, 10 Russian and 10 Thai L1 speakers taken from the speaking subtest of the Finnish Intermediate Level examination. The Finnish Intermediate level certificate is typically used for applying for Finnish citizenship.

The performances were rated by 44 qualified raters in a scale 1-6 for seven criteria: general proficiency, fluency, flexibility, coherence and cohesion, vocabulary, pronunciation and phonological control, and grammatical accuracy (aligning with the CEFR). In addition to this, the raters were also asked to write down their assumptions about the speakers' L1.

In the presentation, we will discuss the results which show that the raters assessed differently the various L1 groups as well as the criteria and the (mis)recognition of speakers' L1, for example the pronunciation of Arabic speakers and the grammatical accuracy and vocabulary of Estonian speakers were rated positively than the other criteria. Previously, based on our analyses, Ahola and Halonen (2021) have discussed assessment of the Thai L1 examinees.

Tuija Hirvelä has a Licentiate in Statistics and a MSc in Mathematics. She has worked as a researcher and a statistician in the Finnish National Certificates of Language Proficiency and recently she has been involved in a research project Broken Finnish: Accent perceptions in societal gatekeeping. Her research interest is

statistical language proficiency assessment with special focus on the item response theory.

Reeta Neittaanmäki is a project researcher and statistician in the Finnish National Certificates of Language Proficiency at the University of Jyväskylä. Presently, she works also as a researcher in project Broken Finnish: Accent perceptions in societal gatekeeping funded by Academy of Finland. Her interest areas are quantitative research methods in the field of language testing with special focus on the Rasch models. She is especially interested in rater effects on performance assessment and rater feedback.

The development and reception of a digital task-based exam for young language learners [D]

Anne van Asseldonk and Lies Strobbe

The Certificate Dutch as a Foreign Language (CNaVT) is currently developing and piloting its first digital, task-based exam for teenagers aged 12-16. The process of development started with a feasibility study that mapped the purpose of the new exam, the intended CEFR-level, technological possibilities and constraints in relation to the test construct, and to the content and skills to be tested. A needs analysis was carried out among potential test takers and other stakeholders such as parents, language teachers, assessment experts and policy makers.

This presentation briefly reports on the feasibility and needs analysis that resulted in the identification of task types to be included in the new digital exam. The presentation elaborates in more detail on the pilot test administered to 20 intended candidates. This sample reflected the variety and geographical dispersion expected in the target test takers, including teenagers from expat communities and pupils attending bilingual education. We report on the follow-up interviews and focus group discussions that were conducted to ensure the appropriateness of the test for the intended young language learners. Areas of focus included, for instance, the skills and strategies needed in the target language use domain, topics of interest, task length and attention span. The answers of the pilot candidates were contrasted with the content represented in the new exam and the formats of representation. In this way, these interviews and discussions deepened our understanding of the piloting data and the learners' experiences with the new digital format and task types. The results of this study shed light on which task types and digital formats to retain, improve or reject, thus

increasing the relevance and acceptability of the new exam to the intended test takers.

Anne van Asseldonk obtained her bachelor's and master's degree in Linguistics from the Radboud University in Nijmegen, the Netherlands. During her studies, she specialized in language testing and Dutch as a second language. She started working for the CNaVT in May 2021 and focusses on setting up a youth exam, but is also involved with the CNaVT exams for adults.

Lies Strobbe is a postdoctoral researcher at the Center for Language and Education (University of Leuven), where she focuses on themes such as multilingualism, language policy, implementation processes in educational innovation and language testing. She has extensive experience as a test developer and researcher at the Dutch as a Foreign Language Certificate (CNaVT). At CNaVT she develops test tasks for different CEFR levels of language proficiency, trains assessors and examiners, performs analyses. In addition, Lies works as a lecturer in the Master of Language Education program.

Multidimensional analysis of syntactic complexity of L2 students' exam writing at the B2+ level [F]

Karlygash Adamova

The present research investigates the way of distinguishing syntactic complexity of L2 English writing by analyzing exam papers at the Basic English Language Examination (BLE). BLE, which stands as an obligatory assessment procedure for English majors at the Hungarian University, is administered to determine the B2+ level of language proficiency (CEFR, 2001).

To this end, 67 grammatical variables were gauged to be measured with the implementation of the Multidimensional analysis tagger (Nini, 2019) and one variable – syntactic structure similarity with the help of Coh-Metrix 3.0 (Graesser et al. 2011). The corpus for the current study was created from 73 texts produced by the students at the written text production part of the BLE.

The full set of linguistic features was classified into sixteen sub-groups adopting Biber's (1988, p. 73-75) method of allocation within grammatical and functional categories. The data analysis process involves the correlation and factor analysis with the implementation of the SPSS package. The correlation was performed conjointly with the range of diversity and co-occurrence of the syntactic

complexity patterns and the grammar points. The factor analysis, on the other hand, adapts Biber's (1988) multidimensional analysis and represents factor loadings to describe alignment of grammatical features in the texts.

Altogether, the piloting study reviews the factors that determine high proficiency L2 writing and attempts to identify the predictive patterns of L2 writing quality with the implementation of computational tools. Taken together, the study seeks to investigate the BLE exam validation process and thus, offers important insights into the field of language testing and assessment.

References 1. Biber, D. (1988). Variation across speech and writing. Cambridge University Press. 2. Biber, D., & Gray, B., & Staples, S. (2016). Predicting Patterns of Grammatical Complexity Across Language Exam Task Types and Proficiency Levels. Applied Linguistics, 37 (5). 3. Council of Europe (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: Cambridge University Press. Online: <https://rm.coe.int/16802fc1bf> 4. Graesser, C., & McNamara, D., & Kulikowich M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. Educational Researcher, 40 (5). 5. Nini, A. (2019). The Multi-Dimensional Analysis Tagger. In Berber Sardinha, T. & Veirano Pinto M. (eds), Multi-Dimensional Analysis: Research Methods and Current Issues, 67-94, London; New York: Bloomsbury Academic.

Karlygash Adamova is a PhD candidate in Applied Linguistics at Pazmany Peter Catholic University (Hungary, Budapest), a young scholar, researcher, Stipendium Hungaricum scholarship awardee and a multilingual person. Her research interest is in the field of English writing, language testing and assessment. Recently, her work is focused on the computer-based and manual analysis of syntactic complexity of English writing aimed to contribute to the validation process of the English examination.

A pilot material for a fair and accessible A2 listening test for adult immigrants with diverse educational backgrounds [DI]

Elina Stordell

Teachers often give us test developers feedback that according to their experience, results from listening tests do not correlate with their classroom assessment. In proficiency tests designed for

Finnish integration training for immigrants, quite many test takers reach lower levels in listening than in speaking, when one could expect quite similar results. Listening scores are also lower than in reading tests.

In the integration training, students have diverse educational backgrounds. Many are less educated and less experienced in formal language tests. They can be able to handle familiar situations in L2 face-to-face in their everyday life, but formal listening tests, which traditionally require also reading and very often writing, can be a challenge.

Test takers with a lower literacy level and possibly with a literacy training background deserve a fair opportunity to show their oral skills which are usually their strongest ones.

Thus, an attempt for a more accessible listening test for A2 level has been made. The test consists of oral and written listening tasks. The oral part is completely auditive and simulates a kind of interview or informal discussion; the written part includes simple multiple choice and short answer questions.

First test versions have been piloted in the integration training including students with various educational backgrounds. This empirical research compares the results from the oral and written listening tasks in one listening test amongst different test taker groups. The presentation reports also test takers views on different task types.

Elina Stordell works as a language assessment specialist and a teacher trainer in Testipiste, a second language assessment centre for adult migrants located in Helsinki. She has experience in item writing, statistical analyses and standard setting, and she is currently interested in adaptive testing.

La evaluación de tareas de revisión en entornos multilingües [F]

Jorge J. Sánchez Iglesias

En la descripción del proceso de escritura, la revisión constituye la fase final y el autor la realiza en su propio texto. Pero en otra perspectiva, la revisión solo “ocurre” cuando la realiza otra persona. Aunque las prácticas de este tipo no son desconocidas en la enseñanza-aprendizaje de lenguas adicionales (segundas o extranjeras), son mucho menos habituales que en L1. Es fácil suponer que ocurre así, en buena medida, porque su foco es casi exclusivamente lingüístico, y quien revisa tiene una responsabilidad

limitada en términos del uso transaccional y/o creativo que en el texto se realiza. De hecho, en el MCER apenas encontramos menciones a dicha forma de elaboración, ni siquiera en el caso de los niveles superiores. En el Volumen complementario, por su parte, aparecen variadas menciones de actividades que incorporan la “revisión”, u otras cercanas como “reelaboración” o “edición”.

Esta propuesta surge en la intersección de dos circunstancias: por una parte, cada vez es más común encontrarse con entornos multilingües de enseñanza, en los que se suman alumnos nativos y no nativos; por otra parte, la revisión constituye una actividad determinante en algunos contextos, como puede ser la traducción, donde además está adquiriendo cada vez mayor importancia merced al desarrollo del componente tecnológico.

Nuestro objetivo es analizar las dificultades que plantea la evaluación de prácticas de revisión realizadas por aprendices no nativos en el marco de un máster en traducción. Definido como un entorno multilingüe, abordaremos dos cuestiones: en primer lugar, cómo afecta la condición de nativo/no nativo a la realización de dichas tareas, considerado su íntima vinculación con el dominio de la lengua; y, en segundo lugar, si para la evaluación de dichas prácticas, por su objetivo, resulta más adecuado el empleo de una valoración global o analítica tanto en el caso de unos como de otros.

Jorge J. Sánchez Iglesias (PhD) read Spanish and Italian at the University of Salamanca (Spain). He is a lecturer of Spanish and Discourse Analysis at the School of Translation and Documentation (Library Science). He previously taught Spanish as a foreign language at Cursos Internacionales. His research focus on assessment and reading/writing interfaces, considered from an applied perspective (first/second language learning, translation).

Piloting Automatic Computerized Testing to Evaluate the Language Abilities of Pre-service World Language Teachers [D]

Michael Vrooman

In order to be certified to teach world languages in most of the US, teachers must achieve the Advanced-Low level on the American Council on the Teaching of Foreign Languages Oral Proficiency Interview (ACTFL OPI). However, apart from having a mock interview with a trained OPI tester, students often do not know how to gauge their oral proficiency before taking the OPI.

This presentation reviews the implications of a pilot study that assesses the language proficiency of pre-service world language teachers using an objective and less costly test that has been found to have a strong correlation with the OPI (Pearson Education, 2018). The Versant Spanish and French tests are automatically scored and cost efficient. The tests employ automatic speech recognition and measure the latency of the response, the rate of speaking, the position and length of pauses, the stress and segmental forms of the words, and the pronunciation of the segments. The results of the test provide the test takers and their instructors with meaningful information regarding the students' language abilities in a timely manner to allow for individualized feedback before the students take the high-cost and high-stakes OPI.

Pre-service teachers at a mid-sized university took the Versant Test and completed a survey to elicit information regarding their experience taking the test and their opinions of its usefulness. This presentation will share those results and insights regarding the use of this assessment to gauge pre-service teachers' language proficiency and their ability to successfully prepare for the OPI.

Additional Authors: *Bret Linford and Carmen Fernández-Flórez.*

Michael Vrooman obtained his PhD in Hispanic Linguistics from the University of Amherst, Massachusetts. He is currently an Associate Professor of Spanish at Grand Valley State University in Allendale, Michigan, where he has taught a wide range of Spanish language, linguistics, and language teaching methodology courses for twenty years. His scholarly work includes publications and workshops on language teaching and learning, language proficiency, and virtual exchange.

A flexible framework: matching student assessments to the CEFR descriptors in a hybrid context [F]

Steve Issitt

Despite the effects of two years of pandemic, when conventional educational practices were often different, university life continued. Staff responsible for intensive language programmes were obliged to rapidly alter their modes of delivery while maintaining the efficacy and accuracy of their assessment methods, not least in reference to established competency criteria such as the CEFR. It may now be anticipated that universities will revert to their pre pandemic model of face-to-face provision, albeit with some exceptions. This

presentation examines the issues related to matching student assessments in these specific circumstances to CEFR descriptors and contingent external English language tests such as IELTS. I outline how this was achieved at the University of Birmingham, how it is currently practised and how it is likely to be implemented in the future.

A key focus of the discussion will be the production of transparent, user-friendly marking criteria which are congruent with the CEFR and adapted to the requirements of our intensive pre-departmental entry EAP programme. I also consider the relationship of the framework (the CEFR) to the students, citing their views of perceived and actual on-going competencies. I move on to consider how a university may benefit from a programme of provision which aligns itself to an internationally recognised benchmark and then discuss how this alignment is crucial to maintenance of the UK's position as leader in the global international student market. Related issues which are also addressed include the development of student confidence and the avoidance of plagiarism, the importance of research-led inhouse assessment materials and the exigencies of practical test construction. Overall, by making reference to provable, consistently improving student performance levels, I present an optimistic perspective and consider that with sufficient sensitivity, student assessments can be framed effectively and meaningfully in changing educational and social contexts.

Steve Issitt works at the University of Birmingham where he is involved with the provision of EAP programmes for international postgraduate students. He is especially interested in evaluating the effects of EAP courses on writing development and holds a PhD in applied linguistics.

Using multi-level tests in benchmarking projects in Iberia [D]

David Bradshaw and Victoria Peña Jaenes

Language skills have been considered crucial in societies where citizens travel, study and work abroad. In this context, language programmes have been implemented at national, regional and school level to improve students' foreign language ability. Assessment tools are considered valuable to offer diagnostic information and to measure advances in proficiency levels. Students' performance in tests can be used to help evaluate and measure the efficacy of English language education systems; to inform decisions on how to improve these programmes, establish target exit levels and measure progress towards these, and to inform the choice

of learning materials and teacher training courses. Cambridge University Press & Assessment has developed Multi-Level Tests (MLT) to offer an assessment solution to support language programmes. The tests include multi-level computer-adaptive Listening and Reading components, and multi-level computer-based Speaking and Writing components. Several versions are available to meet the needs of different age groups of test takers and it has been designed to measure ability and report it in levels aligned to the Common European Framework of Reference for Languages (Council of Europe, 2001). The tests utilise automarking for Listening, Reading and Writing while the Speaking component is examiner-marked.

This presentation will focus on the use of the MLT tests in several benchmarking projects at regional and at school group level in Spain. The MLT solution measured the English language proficiency levels of Primary and Secondary students, and the results informed the recommendations to capitalise on existing strengths and mitigate weaknesses of these programmes. Throughout the session the main features that make the test a suitable option to assess language programmes will be discussed. Finally, reflections about the use of the test will be shared and opportunities for future research will be identified.

David Bradshaw has worked for Cambridge University Press and Assessment for eight years and is currently Senior Assessment Services and Recognition Manager for Western Europe, Iberia and Italy, based in the Madrid office. He has been responsible for the delivery of several large-scale assessment projects across the region designed to measure the effectiveness of local English language programmes within the school system. He holds an M.Ed. with Applied Linguistics from the Open University.

Victoria Peña has worked for Cambridge since September 2018. Prior to that, she taught English to preschool, primary and secondary school students, as well as preparing adults for Cambridge examinations. Since joining Cambridge, she has been involved in creating content for teachers, collaborating in different assessment projects and has delivered presentations and webinars for teachers. Victoria holds a degree in Translation and Interpreting and an MA in Applied Linguistics and a PhD focusing on the washback of proficiency exams.

Implementing a communicative classroom-based assessment program in Dutch lower-form secondary education: teacher and student perceptions [F]

Charline Rouffet

In foreign language (FL) education, the main objective is to learn to communicate in a FL. However, despite the fact that the learning objectives of most language curricula are nowadays based on the CEFR, they are not always implemented well in teaching and assessment practices. The reasons for this can be practical (lack of materials) but also conceptual (lack of expertise) and may contribute to a lack of alignment in language curricula. Under the principle of constructive alignment, assessment and learning activities should be in line with learning objectives in order to enable effective teaching. Evaluation particularly plays a key role, because it has a direct impact on teachers' pedagogical choices and students' learning behavior (i.e., washback-effect). It is therefore crucial to develop and implement (summative and formative) assessment activities measuring the ability to communicate in a FL. However, this is a complex task for teachers due to a lack of suitable material available and a thorough knowledge of assessment methods and their application in practice. Therefore, in our design-based research project, FL teachers worked in co-design with a researcher (i.e., the first author) to create tools (criteria, frameworks and examples) for the implementation of a communicative classroom-based assessment program with a combination of formative and summative functions. The current paper reports on how the co-created materials were implemented in Dutch lower-form secondary education in various schools and on how teachers and students experienced their implementation. Through the analysis of assessment materials and classroom observations, we first observed whether the implementation was carried out as intended. Subsequently, teachers' and students' experiences were investigated through logbooks, a survey and interviews. In this presentation, we will focus specifically on how teachers and students perceived the implementation in terms of practicality and efficiency.

Additional Authors: Prof. Dr. Rick de Graaff, Utrecht University and Dr. Catherine van Beuningen, Amsterdam University of applied sciences

Charline Rouffet is a French teacher trainer at Utrecht University of applied sciences. She also works as a researcher within the research group Multilingualism and Education at the expertise center of the same university. She is currently conducting a PhD research project on the effect of classroom-based assessments on teachers' pedagogical approaches.

Using Screencast Feedback to Improve Undergraduate Students' Speaking Skills in English [D]

Yéssica Aguilera Fuentes

Being able to communicate effectively in English is a skill required to succeed in the academic and workplace settings. However, speaking is one of the most difficult skills to develop when learning a new language, particularly English. In this paper, we present an innovative tool used to provide feedback to students' speaking performance in the context of teaching English as a foreign language (EFL). For this purpose, a screencast feedback technique was used with undergraduate students from a Chilean university to provide feedback on student's individual performance when giving monologues in English. Preliminary findings suggest that the use of screencast feedback has a positive impact on students' academic achievement as it helps them improve their speaking performance being grammar and pronunciation the linguistic components that benefit the most. In addition, screencast feedback is highly valued by students who see it as a motivating and promising strategy.

Yéssica Aguilera Fuentes holds a Master's degree in Curriculum and Assessment and is currently studying a Doctorate degree in Education at a Chilean university. She is an English language teacher with 16 years of experience teaching the language to students of different ages. During the last years she has participated in different research projects dealing with formative and summative assessment in the context of teaching English as foreign language.

Practical model of intensive reading assessment with ReadTheory in the Reading Club [D]

Salwa Said Khalfan Al Anqoudi

ELT teachers found it difficult to engage students in online Reading classes during the Covid-19 Pandemic. The study's primary goal is to increase reading motivation among foundation students by supporting and engaging them in intensive reading activities using the ReadTheory website in the Reading Club Program. Six foundation students registered in the Reading Club through the ANJIZ Center's Programs at the University of Nizwa participated in a six-month action research. The students volunteered to participate in the Reading Club Program, which used the ReadTheory website to teach reading skills. Replicating intensive reading with ReadTheory enabled tracking progress over time in order to improve intensive reading with ReadTheory. Online on-task engagement and performance observations were triangulated with semi-structured one-on-one interviews with each participant as part of the cyclical iterative data collection and analysis. The interpretation derived from thematic coding analysis was enhanced by triangulation across multiple instruments. The findings illustrate the positive impact of online assessment in foundation learners' reading motivation in terms of tracking performance reports and immediate feedback. The intensive reading distance program implemented by ReadTheory resulted in a positive attitude and sustained the learners' long-term intrinsic and extrinsic motivation to read in English. The resulting Intensive Reading Distant Program, which includes ReadTheory, will benefit teachers, students, and all stakeholders interested in encouraging EFL students to read in the target language.

This presentation will benefit ELT teachers and educational stakeholders who want to use e-assessment learning resources to teach Reading skills in the classroom. Thematic coding analysis was enriched by triangulation across multiple instruments.

Additional Authors: *Cecile Aline Gabarre, Serge Gabarre and Siham Mohammed Saleh Al Aamri.*

Salwa Al Anqoudi is a supervisor of English Language Enhancement Center called (ANJIZ) and TESOL instructor with the University of Nizwa, Oman. Her fields of interests include foreign language learning and acquisition, educational technology, engaging environments, mobile learning and qualitative research.

El Manual oficial del SIELE y su integración con la Extensión digital del examen [D]

Juan Miguel Prieto Hernández

Para tener éxito en un examen de dominio como el del Servicio Internacional de Evaluación de la Lengua Española (SIELE) es fundamental conocer las tareas de las pruebas del examen. Con el fin de informar tanto a candidatos como a preparadores, presentamos un manual en el que, por medio de un método semi-inductivo, se pone el foco primero en la práctica y, posteriormente, en los contenidos explícitos. El trabajo práctico que se realiza permite analizar y conocer algunas de las estrategias que se utilizan para facilitar la resolución de las tareas: anticipación del contenido, identificación de la información relevante de cada ítem o modo en que se discrimina la información correcta en cada caso.

Esta profundización en el conocimiento de las tareas resulta relevante, también, por el efecto rebote de los exámenes en el aprendizaje de lenguas. En ese sentido, se puede observar la peculiaridad de los exámenes multinivel como SIELE, en que los candidatos tienen niveles de competencia muy diferentes. Dentro de ese contexto, el manual de preparación del SIELE se puede mostrar como una solución práctica que ayuda a la preparación del examen y a llevar al formato en papel las consecuencias de un examen en línea.

Resulta novedosa, a este respecto, la existencia de una Extensión digital con diferentes secciones en las que los candidatos podrán escuchar los audios de las pruebas, contestar a las preguntas, revisar sus respuestas y comprobar sus aciertos, y repetir las tareas y las pruebas todas las veces que quieran. En las pruebas de EIE y EIO encontrará, además, información sobre "cómo se califica" cada tarea, con ejemplos de producciones auténticas de candidatos. De este modo, el usuario del manual podrá realizar un entrenamiento real con los modelos del SIELE en la misma plataforma en la que se administra el examen.

Additional Authors: *Inés Soria Pastor, Instituto Cervantes, Georgina Lacanna, Universidad de Buenos Aires y Luis Miguel Samperio Sánchez.*

Juan Miguel Prieto holds a PhD in Spanish Philology and is currently in charge of the Assessment Department of Cursos Internacionales de la Universidad de Salamanca. He represents the University of Salamanca in The Association of Language Testers in Europe (ALTE). He is the author of many specialized articles on the methodology

of Spanish teaching and the assessment of Spanish as a foreign language. Additionally, he has contributed his expertise to several teaching materials as a co-author.

Online language testing: investigating the candidates' attitudes and reactions [D]

Letizia Cinganotto

The presentation aims at describing the ongoing research carried out at the Centre for Assessment and Language Certificates (CVCL) at the University for Foreigners of Perugia, Italy related to the digital evolution of assessment and training, after the COVID-19 pandemic.

The main features of the different tests and certificates provided by the CVCL will be addressed as a starting point, in particular the Italian language certificates named CELI, aligned with the CEFR and the methodological certificate named DILS. These were developed for Italian language teachers aiming at developing methodological competences in the teaching of Italian L2/FL.

During the pandemic a wide range of online training initiatives were organized in order to prepare teachers and students for the different types of tests. Lessons learnt from the emergency situation are being capitalized in order to possibly reshape and improve language testing and assessment procedures, also taking into account the literature in the field.

The presentation will highlight preliminary results of a study conducted with a sample of teachers and students aimed at investigating their reactions and attitudes towards different dimensions related to online testing, such as timing, the use of the device, their level of anxiety, their digital literacy etc. Interviews and focus groups, as main qualitative methods, have been designed in order to collect useful information for further developments, in preparation for the possible future digitalization and innovation of the entire process of Italian language testing at CVCL.

The research begins with comparisons with “pen and paper” version of the tests, in order to collect the participants' reactions and to get the best out of the traditional process in order to build innovative digital assessment pathways. Main categories grouping the participants' comments collected through the focus groups and the interviews are described and commented on according to the Framework Analysis.

Letizia Cinganotto, former Senior Researcher at INDIRE (National Institute for Documentation, Innovation and Educational Research), Italy, currently teaches language teaching at University for Foreigners of Perugia, Italy. She holds a PhD in synchronic, diachronic and applied Linguistics. She is a member of different working groups and scientific committees on CLIL and language learning both at national and international level. She has presented papers at national and international conferences and published articles and chapters in peer reviewed journals and five volumes on CLIL.

Mapping the SMEEA Gaokao tests to the CEFR [F]

Jane Lloyd

This paper reports on a project where the Shanghai Municipal Educational Examinations Authority (SMEEA) engaged Cambridge University Press and Assessment (Cambridge) to begin the alignment process of their six foreign language tests to the Common European Framework of Reference (CEFR) (Council of Europe, 2001, 2020). These six language tests (English, French, German, Spanish, Russian and Japanese) are developed by SMEEA and form part of China's National College Entrance Examination system. They are referred to as China's National College Entrance Examination (Shanghai Paper), or the Shanghai Gaokao. This project was undertaken to enable a comparison of the relative difficulty of each language version of the Shanghai Gaokao, in terms of the CEFR. Additional project aims were to carry out training in the CEFR and in CEFR mapping procedures, using a blend of online delivery in English supported by materials in Chinese, and a cascaded training model which was delivered in Chinese and run by dedicated trainers for each foreign language team. This project took place between September 2021 and January 2022, involving a series of linked training and mapping activities and workshops. The format of the activities in the project followed the broad lines as set out in the Council of Europe's Manual for Relating Examinations to the CEFR (2009) and its updated Aligning Language Education with the CEFR: A Handbook (2022). In this paper the focus is on the practical aspects of a computer mediated multilingual CEFR mapping project, on our reflections and recommendations, and on the participant feedback.

Additional Authors: *Graham Seed, Cambridge University Press and Assessment, Xu Wen, Shanghai Municipal Educational Examinations Authority, Hu Yue, Shanghai Municipal Educational Examinations Authority.*

Jane Lloyd is a Senior Research Manager at Cambridge University Press and Assessment. Before joining Cambridge, Jane worked in Japan, Southeast Asia, Turkey and the Middle East as a teacher, trainer and manager in state schools and higher education for over 20 years. She has a degree in Linguistics, a DELTA, an MA in Linguistics and TESOL, and an MA in Language Testing. She is currently studying for a PhD in Language Testing at CRELLA.

Italian language testing regime: alternative perspectives [F]

Paola Masillo, Giulia Peri and Sabrina Machetti

Over the past two decades, a significant number of European countries have introduced linguistic requirements for the purposes of migration (Rocca et al., 2020). In Italy, to obtain a long-term residency permit non-EU citizens must demonstrate a certain level of language proficiency in L2 Italian (Law no. 94/2009). One of the options to demonstrate this proficiency is successfully passing an Italian language proficiency test (M.D. 4th June 2010). During the past two decades, the social, cultural, and economic context has undergone great changes (technologies used, cultures and languages involved in the everyday practices). Therefore, the competencies required of anyone living integrated in the dynamics of a globalized, diversified, and increasingly technological world have changed (Purpura, 2021). When reflecting on issues of integration and language policies, it is also necessary to reflect on the nature of competencies required and on the tools used, to understand whether they are actually adequate to reflect the real-life behavior expected of migrants in the host country. In order to do so, the paper reports language testing practices carried out at national level within the framework of current legislation, focusing on the use of language requirements in order to issue a long stay permit for non-EU citizens.

The study analyses the impact of those language policies to migrants and is linked to academic research that investigates the language test as a power tool (Shohamy, 2011). Therefore, the paper aims at (a) reporting the development of policies and practices designed to support the linguistic integration of adult migrants in Italy (Machetti et al., 2018) and (b) exploring new perspectives for this purpose. For this purpose, we advocate the adoption of a Learning-Oriented Approach to Assessment (Purpura & Turner, 2018) by proposing the development of a scenario-based test for L2 Italian (Purpura, 2021). Both the approach and the technique propose an alternative to traditional, decontextualized testing with distinct skills, emphasizing the dimension of technology, peer

interaction, the importance of feedback and assessment also as a process aimed at learning.

Paola Masillo is PhD in “Linguistics and Teaching Italian as second language”, University for Foreigners of Siena. She graduated at the University for Foreigners of Siena in Linguistics and Intercultural Communication, with a dissertation on Language Testing and Assessment. Her main research focuses on learning Italian as a foreign/second language, language assessment and language policies. She has been a post-doctoral research fellow at the CILS Centre of the University for Foreigners of Siena, where she is currently working in the technical-scientific and data processing area.

Giulia Peri is a PhD candidate in Applied Linguistics at the University for Foreigners of Siena. Her research interests focus on L2 Italian Teaching and Learning, Language Testing and Assessment and Language Testing and Technology. She has worked as a Research Fellow at CILS Centre - Certification of Italian as a Foreign Language of the University for Foreigners of Siena. She has also worked as a Language Assistant for Italian and English in the Canary Islands and Australia.

Sabrina Machetti is Associate Professor in Modern Language Education at the University for Foreigners of Siena, where she obtained her PhD in Linguistics and Teaching Italian to Foreigners. She teaches courses in Language Testing and Assessment. In 2018, she was a visiting scholar at Teachers College, Columbia University. She is currently the director of the CILS Certification Centre of the University for Foreigners of Siena. She has written articles in several Applied Linguistics journals and edited volumes.

Teaching the teachers: Designing digital assessment for language teachers which both evaluates and educates [D]

Nicole Busby and Anja Angelsen

This study describes student responses to a new digitally based formative assessment design in a linguistics course for in-service language teachers in Norway. The goals of the redesign were to increase flexibility by transitioning to a fully digital format and to increase the relevance and validity of the assessment, as well as introducing assessment approaches that the students could

implement in their own teaching. A major challenge in adapting to unsupervised online assessment is ensuring test validity, reliability and credibility. It requires a shift in focus from assessing memorised knowledge to assessing understanding of concepts and the ability to apply this understanding to new examples. We created an assessment package that integrated formative and summative assessment: the summative component built on continuous work and formative assessment and included student reflections about the learning process. Online surveys were used to collect data about students' perceptions of the assessment as part of investigating the validity of the new design. Students reported finding the assessment design relevant for their own learning and as a model which could be used in improving their own teaching. They reported increased motivation and greater understanding resulting from being able to work on their assignments following feedback and having smaller tasks throughout the semester enabled them to more evenly distribute their workload. Students also reported feeling that the combination of formative and summative assessment gave them greater opportunities to demonstrate what they had learned from the course. This assessment package appears to be a useful design that reflects real-world situations and may also be suitable for other types of courses as an alternative to supervised exams, online or in person.

Nicole Busby is an associate professor in the Department of Language and Literature at the Norwegian University of Science and Technology in Trondheim. She teaches online and face-to-face courses in English linguistics, language acquisition and research methods. Her research interests include academic reading, vocabulary, and second language acquisition.

Anja Angelsen is an assistant professor in the Department of Language and Literature at the Norwegian University of Science and Technology in Trondheim. She teaches online and face-to-face courses in English language proficiency, linguistics, and translation. She coordinates the continuing programmes at her department and has considerable experience of teaching courses for in-service teachers. Her academic interests include translation, vocabulary, and language learning and teaching.

Assessing L2 Speaking Performance Asynchronously and Online: On Navigating the Fields of Tension between Authenticity, Validity, Reliability and Practicality [D]

Olivia Rütli-Joy

Amidst the wake of the COVID-19 pandemic, exponential technological developments and unprecedented global change have profoundly affected the way languages are learned and taught. Consequently, there is a pressing need to reconceptualise language assessment and rethink quality criteria such as authenticity and validity in the digital age. This paper presents an LSP test developed to assess Teacher Language Competence in English and French in a digital environment. The competence-oriented L2 oral performance test is administered asynchronously via an LMS and contains seven integrated video-based test tasks. The video-vignettes depict prototypical L2 classroom situations filmed from the first-person-perspective to establish a close relationship to the target language use (TLU) domain and prompt the test takers to produce spontaneous test responses in the target language, which they record directly in the LMS.

To identify the affordances and limitations of this test, 415 test task responses in the target language English were rated by four independent raters and rater bias and interaction analyses were conducted by means of an MFRA and a partial credit model. The analyses revealed significant differential rater functioning and inconsistent rater severity and leniency overall. Furthermore, stark variability in terms of rater-task interaction per individual test task was revealed, indicating that there was almost no uniformity maintained within the ratings across raters and tasks. Based on these findings, the author discusses the affordances and challenges of online-mediated, vignette-based speaking assessment related to authenticity, validity, reliability and practicality. Furthermore, challenges of navigating various fields of tensions are discussed. Such fields of tension include, among others, offering a broad enough range of authentic tasks to elicit skills relevant to the test taker while maintaining comparability of test-task interpretation and test-task responses, or ensuring high ecological validity in an asynchronous test-setting that seeks to replicate the TLU domain in a virtual environment.

Olivia Rütli-Joy is a senior research assistant at the University of Fribourg (Switzerland). Her background is in English and Media

Studies (BA), English Literature (MA) and in English language teaching at lower and upper secondary school as well as tertiary education in Europe and New Zealand. She received her PhD in foreign language teaching and learning. Her research interests are in foreign language teaching and learning, language / LSP testing, educational assessment, teacher education, and feedback processes.

Interpret and integrate ideas and information in PIRLS [D]

Lena Edander

In the study, reading comprehension is examined in student-produced answers to tasks of the type Interpret and integrate ideas and information in the international reading comprehension test PIRLS. The purpose is to highlight how reading comprehension is expressed in student-produced answers to tasks within the said reading process when the student managed to get full points, respectively when the answer only gives one point. The study is carried out to get in-depth information about how Swedish students in year 4 succeed in answering questions where they are expected to formulate answers that require their own interpretation. The study also provides information on how extensive answers in terms of written formulation and content are required to answer tasks within the reading process Interpret and integrate ideas and information. The relevance of the didactic grant is based on results from previous research that showed that Swedish schools support students' basic reading skills, but that in-depth reading comprehension is not catered for to the same degree (Liberg 2010). The material for the study consists of all Swedish results on two questions in PIRLS 2021 where the maximum score is three. The method includes categorization of question types, an analysis of the assessment instructions regarding tasks of the nature of Interpreting and integrating ideas and information, and a categorization of the student answers as concise or elaborated. Then a substantive and linguistic analysis of student-produced answers with full points in relation to answers with one point has been done. In summary, the tentative results show interesting differences both regarding the substantive parts of the answers as well as the scope of the answers.

Lena Edander is a doctoral student at the department of Education at the university of Uppsala, Sweden. Her research interest is reading literacy and how this is measured in large scale testing. Text analysis of students answers combined with quantitative data from PIRLS

is used in her research to reveal how students handle advanced reading comprehension items and to find ways to improve reading comprehension.

Marcadores y certificación lingüística [D]

Marta García

En esta comunicación nos proponemos mostrar la importancia de los marcadores discursivos en el mundo de la Evaluación y la Certificación lingüística del español. Los candidatos que deciden certificar su nivel de español deben resolver tareas de expresión e interacción escritas y orales en las diferentes pruebas que conforman los exámenes, y es ahí donde los marcadores tienen su protagonismo. Por eso, vamos a centrarnos, por un lado, en las tareas de Expresión e interacción escritas que aparecen en diferentes exámenes de español y en sus escalas de corrección. Por otro lado, vamos a analizar producciones escritas de algunos candidatos y la actuación de los calificadores de esas pruebas. Con la recogida de estos datos, podremos llegar a conclusiones que para un profesor de español pueden ser muy útiles a la hora de ayudar a los candidatos de certificaciones lingüísticas a preparar pruebas de Expresión e interacción escritas y también para enfocar el contenido de sus clases de español.

Marta García holds a B.A. in German Studies, a Master's Degree in Teaching Spanish as a Foreign Language and a Doctorate in Spanish Philology. She has worked as a teacher in Cursos Internacionales de la Universidad de Salamanca since 1998. She is responsible for LC USAL esPro test, a test of general Spanish language for professional purposes. In ALTE she has participated in numerous conferences and also in the Study on Comparability of Language Testing in Europe (DG, EAC, European Commission) as Content Analyst and in the project Surveylang. She is part of and has contributed to the working group LSP, producing and reviewing draft texts for the Guidelines for the Development of Language for Specific Purposes and is a member of the Standing Committee.

“No one gets left behind”- Getting stakeholders ready for the digital age [D]

Isabel KnöB

Over the last two years, the digitalization of language assessment has accelerated at an extraordinary speed. While it has been amazing to see how quickly language assessment providers have adapted, we should not forget that digitalization does not stop at simply developing a digital examination or launching software for language assessment. Recent technological evolutions have led to a variety of new challenges and have uncovered insecurities regarding both the process of delivery as well as the validity of digital assessment vis-à-vis traditional pen & paper assessment for all users “behind the screen”.

Traditionally, language teaching and assessment has been a non-technical work environment. Now people involved in language assessment have been forced to adapt and implement these technological evolutions as well as develop additional skills at an immense speed. The ALTE minimum standards highlight the importance of communication with, and inclusion of all stakeholders involved. Besides the assessment provider, the stakeholders include the language learners and test takers, examiners and markers, examination centers, and institutions.

This presentation will demonstrate different strategies, including hands-on tips, communication and training tools and other solutions telc employed to facilitate stakeholders transition through this adaptation process. The presented data was accumulated from an exploratory study carried out during the introduction and implementation of telc digital examinations for examinations outside the EU starting in 2020.

Isabel KnöB is a project manager at telc, where she supervised the introduction of digital exams in 2020. She has a master's degree from the University of Applied Sciences in Karlsruhe, where she focused her research on the preferences regarding user experience across multiple channels with a focus on generational divergence. She specializes in digitalizing different aspects of telc's language assessment product range and is responsible for iteratively modernizing the mediums of paper-based and digital exams.

The implementation of the CEFR in Sweden: the relation to the national curriculum standards for languages [F]

Maria Håkansson Ramberg

Since the publication of the Common European Framework of Reference for languages, CEFR (Council of Europe, 2001), the document has become a widely used framework, not least for defining levels of proficiency. In the field of language learning and teaching, this has resulted in an increased use of the CEFR as a reference point for the alignment of tests, textbooks and curricula. Aligning national curriculum standards to the language levels of the CEFR has been advocated in European countries. In the Swedish educational system, the six CEFR levels are comparable to the seven courses in the national curricula for foreign languages. Although the Swedish syllabus for languages is clearly influenced by the reference levels of the CEFR, no systematic empirical studies of this alignment have been made. Against this background, this present study was designed with the aim of examining the validity of the assessment of students' written proficiency at different courses of the Swedish school system in relation to the CEFR.

The data comprise 60 texts written in L2 German by Swedish secondary school students in courses aiming for three different steps of the national curriculum (roughly comparable to CEFR levels A2.2, B1.1 and B1.2). The essays were scored by (1) Swedish experienced teachers using Swedish national performance standards and (2) trained, external raters using scales from the CEFR. Analysis of scores were based on quantitative and qualitative methods, for example descriptive and correlational statistics and comparisons between rater comments. The findings of the present study confirm that the pass level at the end of secondary school is roughly equivalent to a B1-level. Additionally, certain similarities and differences between the statements of the two groups of raters are presented and the role of the CEFR as an external reference point for national educational standards is discussed.

Maria Håkansson Ramberg obtained her PhD in German Linguistics from Uppsala University, Sweden. Her research interests are German didactics, language assessment, classroom research and teacher education. She is a qualified teacher with almost 15 years of experience from teaching secondary school in Sweden and is currently working as a teacher educator and researcher at Uppsala University.

The ITACE for Lecturers: a locally developed C1 test for lecturers in Flemish higher education [F]

Frank van Splunder, Catherine Verguts and Tom De Moor

This presentation discusses the implementation of a high-stakes and mandatory English language test for lecturers in Flanders, the Dutch-speaking part of Belgium. The test was developed in response to a political decision by the Flemish government, which legislated that lecturers need official proof of the advanced C1 level of the CEFR when teaching in a language other than their mother tongue. In order to test whether lecturers in Flemish higher education have the required C1 level, the ITACE for Lecturers (Interuniversity Test of Academic English for Lecturers) was developed by a consortium of four Flemish university language centres.

While standardized large-scale tests such as IELTS and TOEFL typically cater to an Anglo-American educational context, locally developed tests may be more appropriate in a specific non-English dominant context where English is used as an additional medium of instruction.

In the presentation we illustrate how the ITACE for lecturers was constructed as a digital, domain-specific and purpose-built language test. It was mapped onto the CEFR C1 level, both for the productive tasks and the computer test, which includes vocabulary, grammar, reading and listening. We will discuss some choices made in the construct of the ITACE for lecturers, describe the assessment procedures, and highlight its reliability based on test and item analyses.

While it is a matter of debate whether the C1 level is necessary for teaching purposes, we believe that this advanced level is justified to assure quality in English-taught study programmes. Finally, we discuss how the test has shaped higher education practices by setting up English and EMI courses for lecturers and providing additional support, for instance feedback sessions and individual coaching. As a result, most lecturers now show an increased awareness of the adequacy of their English production, thus enhancing the EMI experience for all parties involved.

Additional Author: Sarah De Paepe (Ghent University).

Frank van Splunder holds a PhD in Applied Linguistics from Lancaster University. He teaches academic writing and he is a language tester at Linguapolis, University of Antwerp. The focus of

his research is English as the language of globalization and its use in higher education, particularly in Flanders and the Netherlands.

Catherine Verguts teaches Dutch and English as a second language at the Ghent University Language Centre. She specialises in specific purposes and academic language skills. Catherine is involved in the development of high-stakes tests of Dutch and English in the Interuniversity testing consortium.

Tom De Moor has been working for the language centre (UCT) of Ghent University since 2008, where he currently coordinates and teaches Business and Academic Language Education. He specializes in academic English for doctoral and professional staff. Among other courses, he has codeveloped and teaches "Lecturing Skills in English", in the context of which he wrote the book "Teaching Through English - The didactics and language of English-Medium Instruction in practice" together with Linda Mous (Tilburg University).

Language Background and its Effect on Performance on a Digital Age Test [D]

Nathan T. Carr

This study aims to investigate the effect of test takers' language backgrounds on their performance on the Duolingo English Test (DET), a truly Digital Age test. The DET is a digital-first computer-adaptive test, and uses human-in-the-loop AI technologies for task auto-construction, automarking, and remote proctoring.

A number of previous studies (e.g., Angoff & Sharon, 1972; Bachman et al., 1995; Kunnan, 1995; Stricker, 2002; Stricker & Rock, 2008) have examined language background effects in the context of other tests. Building on these previous studies, this study poses the following research questions regarding the fitness of this digital test for test takers of various groups:

1. To what extent do L1, L2, and bilingual users of English perform differently on the DET?
2. To what extent do L2 users of English from different countries in terms of prevalence of English use in educational, entertainment and business contexts (i.e., the Outer and Expanding Circles, based on Kachru, 1984) perform differently on the DET?
3. To what extent do L2 users of English from different groups in terms of years of classroom instruction in the English language perform differently?

4. What patterns exist in the relationship between test takers' first language and their dominant academic languages?

This quasi-experimental study addresses these questions by having participants take a language background survey, followed by an operational DET. The survey consists of sections on DET familiarity, participants' use of English as a first or second language, what language the participants consider to be their strongest academic language, and the extent of their study of English. The data are analyzed using MANOVA and factor analytic methods (exploratory and confirmatory factor analyses). The results are discussed in terms of their implications for how test scores serve as appropriate reflections of language ability for different groups.

Nathan Carr is a Professor of TESOL at California State University, Fullerton. He earned his Ph.D. in Applied Linguistics at UCLA, with an emphasis in second language assessment. His research interests are eclectic, focusing primarily on language assessment, particularly validation, computer-automated scoring of short-answer tasks, test development and revision, rating scale development, task characteristics, and assessment literacy training. He has lived and taught in Taiwan and Kazakhstan, and has done assessment and curriculum projects in China, Vietnam, Indonesia, and Azerbaijan.

Paper Presentations

FRIDAY 28th APRIL

Implementation of mediation in Catalan EOI certification exams [F]

Montserrat Montagut Montagut and Josep Maria Montcada Escubairó

This session will report on the procedures followed in the Escoles Oficials d'Idiomes (EOI) in Catalonia to introduce a mediation paper in their language certificate tests for adult learners at four CEFR levels (B1, B2, C1 and C2) and for eight different languages (Arabic, Catalan, English, French, German, Italian, Russian and Spanish).

The publication of the Spanish Royal Decree RD 1/2019 compelled the EOIs in Catalonia and throughout Spain to develop fair and valid mediation tests as part of their high-stakes certification exams.

Also, the publication of the CEFR Companion Volume presented a new challenge in the multilingual context of the EOI. The new scales and their corresponding descriptors provided an operational definition of the new linguistic construct of mediation. In this scenario, our language testing institution had to revisit test specifications and task design in an attempt to tap into these new constructs while preserving test comparability across languages. In addition to this, new rating scales were developed to embody the common constructs. Other issues such as test administration and test marking were also adapted to the new high-stakes testing procedure.

After a brief outline of the context, the impact of academic and non-academic issues arising from this measure on the different steps along the whole testing process will be analysed. Finally, the solutions adopted and the challenges that still lie ahead will be presented.

Montserrat Montagut is the head of the Foreign and Heritage Languages Unit in the Catalan Ministry of Education. She manages educational programs, leads the implementation of Erasmus+ in Catalonia and coordinates the official schools of languages. She graduated in Catalan Philology at the University of Barcelona (1996) and post-graduated in Lexicography and Terminology at Pompeu Fabra University. She previously served at the Catalan Ministry of Culture, from where she led the Assessment Unit for (Catalan official exams), e-learning programs, and digital language tools.

Josep Maria Montcada has been an item writer for high-stakes certification exams administered by the official schools of languages (EOIs) in Catalonia. He trained in language assessment at Lancaster University and has attended numerous seminars on this topic, both as a participant and as a speaker. He is currently responsible for the EOI Unit at the Department d'Educació of the Generalitat de Catalunya, which develops certification tests in 8 languages and 4 levels: B1, B2, C1 and C2.

El efecto de los exámenes de acreditación desde el punto de vista de profesores y alumnos en Jaén (España) [DI]

Victoria Peña Jaenes

Como consecuencia de la relevancia del inglés para acceder al mercado laboral (Chávez Zambano, Saltos Vivas & Saltos Dueñas, 2017:761) y a estudios universitarios en el mundo, la demanda de cursos de inglés y de exámenes de acreditación ha aumentado en las tres últimas décadas en España. A pesar del número de alumnos y de la ya mencionada importancia de esta lengua, muy pocos estudios se han centrado en el efecto que los exámenes de acreditación podrían tener en la enseñanza y el aprendizaje del inglés en este país.

Esta sesión se basa en una investigación más amplia sobre el efecto de los exámenes en el aprendizaje y la enseñanza de idiomas, realizado en Jaén (España), que contó con la participación de 131 estudiantes y 8 profesores. La presentación analizará las opiniones que el profesorado tiene sobre los cursos más enfocados a la preparación de examen, los cursos de inglés más general que imparten y los exámenes de dominio que conocen o para los que preparan a sus alumnos. Estos datos se estudiarán y se contextualizarán con las opiniones expresadas por sus alumnos y las investigaciones más recientes en la materia. Entre otras cuestiones, los docentes identificaron las características y hábitos de su alumnado que pueden contribuir a superar un examen de dominio. Asimismo, reflexionaron sobre cómo afectaba el examen de acreditación a sus clases y a sus alumnos. Tanto profesores como estudiantes describieron los principales desafíos a los que estos últimos se enfrentan en las distintas partes del examen.

El objetivo de esta presentación es entender mejor el efecto que los exámenes de dominio tienen en un contexto muy específico, en el que dichas pruebas gozan de reconocimiento y donde poder demostrar el nivel de inglés es fundamental para mejorar las perspectivas profesionales y académicas.

Chávez-Zambrano, M., Saltos-Vivas, M. A. & Saltos-Dueñas, C. M. (2017). La importancia del aprendizaje y conocimiento del idioma inglés en la enseñanza superior, *Dominio de las Ciencias*, 3: 759-771. DOI: <http://dx.doi.org/10.23857/dom.cien.pocaip.2017.3.mono1.ago.759-771>.

Victoria Peña has worked for Cambridge since September 2018. Prior to that, she taught English to preschool, primary and secondary school students, as well as preparing adults for Cambridge examinations. Since joining Cambridge, she has been involved in creating content for teachers, collaborating in different assessment projects and has delivered presentations and webinars for teachers. Victoria holds a degree in Translation and Interpreting and an MA in Applied Linguistics and a PhD focusing on the washback of proficiency exams.

Overcoming Challenges Aligning Language Assessments to Standards [F]

Fabiana MacMillan, David MacGregor and Jason Kemp

Aligning language tests to external language proficiency standards is increasingly required by policy, often with high-stakes implications. Proficiency at B1 level on the Common European Framework of Reference for Languages (CEFR, 2001) is required for settlement or naturalization in the United Kingdom (Blackledge, 2009). Similarly, the International Civil Aviation Organization sets English language requirements for air traffic controllers and pilots (Alderson, 2009; Papageorgiou, 2016).

In the U.S. K-12 context, federal law requires that states annually assess the “English proficiency of all English learners” (i.e., students formally identified as needing English language support services), with assessments “aligned with the State’s English language proficiency standards” (Every Student Succeeds Act, 2015). The U.S. Department of Education provides a standards and assessments review process detailing how to demonstrate compliance with federal law (U.S. DoE, 2018). Regardless of context, since no one correct way to align a test to standards exists, test developers claiming alignment with standards must defend the consistent application of the selected alignment methodology (Fulcher, 2016).

Webb’s (1999) alignment methodology provides a useful framework for thinking about and planning alignment studies, by defining alignment in terms of categorical concurrence, depth of knowledge, range, and balance, suggesting minimum criteria for each. The associated Web Alignment Tool can help conduct the study and analyze results. However, the methodology and tool were designed for alignment of content-area tests (e.g., in mathematics) to

standards, raising issues in applying it to a language proficiency test. Therefore, per Cizek et al. (2018), we adjusted our criteria for determining alignment given test design and specifications. For example, we relied on Cook’s (2006) guidance tailored to language proficiency assessments.

We discuss challenges faced in using Webb’s alignment framework and how we addressed them to accurately reflect the alignment between the English language development standards framework and the test content.

Fabiana MacMillan is the Director of Test Development at WIDA, University of Wisconsin-Madison, where she oversees the development of a suite of assessments that serve multilingual K-12 learners in the United States and internationally. She received her PhD in Applied Linguistics from the Fluminense Federal University in Rio de Janeiro, Brazil. Dr. MacMillan has published in the areas of language assessment, test validation, and computer adaptive testing.

David MacGregor is an Assessment Researcher at WIDA. He received his PhD in Applied Linguistics from Georgetown University. Prior to coming to WIDA, he worked on various language assessment projects at the Center for Applied Linguistics. He has published articles in journals such as Language Assessment Quarterly, Language Testing, and Studies in Second Language Acquisition. Dr. MacGregor is interested in issues related to test validity, standard setting, and computer adaptive testing.

Jason A. Kemp is an Assessment Researcher at WIDA. He received his PhD in Curriculum and Instruction from the University of Wisconsin-Madison. He participates in projects that support multilingual learners and their educators. He evaluates test items and graphics during bias, sensitivity, and content reviews, and he also conducts qualitative research studies. Dr. Kemp has presented and published in the areas of embedded universal tools, equity in remote testing, and heritage/second language education.

From Mediation to Knowledge Transformation: expanding the construct of the reading into writing task [F]

Alina Reid

The introduction of Mediation in the descriptive scheme of the CEFR reflects the major shift away from the isolated four-skills approach,

to one of skill integration which is becoming ever more prominent in communicative language teaching and assessment. Mediation is particularly relevant to the educational domain, in contexts such as CLIL and EMI, where second language users are expected to interact with and co-construct meaning from multiple sources of information.

This presentation will focus on Mediation in the written mode enacted through the reading-into-writing task type. Like the concept of Mediation itself, the integrated writing task is rapidly gaining in popularity by virtue of its similarity and pertinence to real-life writing tasks that students encounter in the academic domain. When writing for educational purposes, students typically learn new information from external sources, interpret, adapt, and synthesise the new information with their existing knowledge in order to create a new text. They engage in a mutually altering interaction with the sources, which generates a cognitive process of knowledge transformation, also known as the “cognitive trademark” of academic writing (Flower, 1990).

The session will discuss how the CEFR descriptive categories of Mediation can be “mix[ed] and match[ed]” (Council of Europe, 2020), adapted and, indeed, expanded in order to create a reading-into-writing writing task that achieves a fuller representation of writing in the academic domain.

Alina Reid is a Language Test Developer at Trinity College London. She is a highly qualified ELT professional with extensive background in teaching and teacher training in a variety of educational contexts in Spain, Vietnam and the UK. She holds an MA in Language Testing (Lancaster University), and her interests lie in assessing English for Academic Purposes, authenticity in performance-based testing, and rating scale design.

SIELE remoto: el desafío de la evaluación a distancia [D]

Sandra Cruz

La necesidad de adaptarse a las constantes demandas de la evolución de las tecnologías se sumó, a comienzos de 2020, a las exigencias de responder a una situación global inédita que afectó al desarrollo de las actividades en todos los ámbitos profesionales y académicos. El examen SIELE (Servicio Internacional de Evaluación de la Lengua Española), que certifica el grado de dominio del español a través de medios electrónicos, no fue la excepción. Desde sus comienzos en el año 2016, SIELE se propuso como un sistema

de evaluación íntegramente a través de ordenadores que se aplica en una amplia red de centros de examen alrededor del mundo. En el 2020 debió responder a un nuevo desafío, con el diseño de una opción de aplicación y supervisión en modalidad remota que posibilitara la continuidad de la acreditación de los candidatos que no podían desplazarse hasta los centros de examen. Esta nueva modalidad se instaló como alternativa permanente de aquí hacia el futuro. Basado en cuatro pilares fundamentales (fiabilidad, accesibilidad, calidad académica y respaldo institucional), el SIELE remoto constituye una alternativa flexible y moderna que cumple con los estándares de calidad propios de una certificación internacional. Para ello, se han desarrollado medios que permiten la identificación de los candidatos y un seguimiento de las pruebas mediante proctoring, que garantiza que no se cometa fraude, y se ha establecido una nueva relación con los centros de examen, que se siguen ocupando de la administración del SIELE en la modalidad en remoto. El objetivo de esta presentación es describir los principales puntos del desarrollo de esta implementación y compartir los resultados alcanzados hasta el momento.

Additional Authors: *Inés Soria, Juan Miguel Prieto Hernández y Georgina Lacanna.*

Sandra Cruz es licenciada en Psicología y maestra en Lingüística aplicada por la UNAM. Colabora en el CEPE hace 30 años, impartiendo clases de español y en formación de profesores. Es Profesora de tiempo completo y jefa del Departamento de Evaluación, es calificadora del SIELE. Coordinó el nivel 3 de los libros Así Hablamos. Español como lengua extranjera; y Dicho y hecho. Español como lengua extranjera. Fue coordinadora de la Comisión de Enlace de la Academia mexicana de la lengua y consultora en la FUNDEÚ. Creadora y locutora de las cápsulas "Para no eruditors" que se transmitían en el IMER.

An online flipped classroom approach to standard setting: What factors support consistent decision making in online contexts? [D]

Jane Lloyd

This paper reports on how standard setting panellists arrive at their judgements in an online environment, comparing more with less consistent judges. It investigates decision-making processes, factors that influence decisions, resources drawn on, features

attended to, and how these factors and resources relate to high and low levels of consistency in decision making.

The context is operational standard setting for the reading component of an online adaptive test of English as a foreign language for adults, which reports performance from CEFR Pre-A1 to C2. Three standard setting panels were convened online, using a combination of self-access materials and video-conferencing. 21 standard setting panellists took part in the study from locations around the world.

Data consisted of selected response data from a) a questionnaire on factors affecting judgements and decision-making processes, b) open-ended data from individual retrospective reports on decision making on individual reading items, c) group discussions recorded during the standard setting process, and d) 1-1 interviews carried out following the standard setting panels.

Multi-facet Rasch analysis of judgement data was used to identify the most and least consistent panellists. The qualitative data was explored to identify significant differences in the thoughts and behaviours of the two groups. The paper will provide a summary of the statistical data, and categories and quotes from the qualitative data. Implications for operational standard setting include a protocol to support more consistent decision making and a robust methodology for investigating the decision-making process.

Jane Lloyd is a Senior Research Manager at Cambridge University Press and Assessment. Before joining Cambridge, Jane worked in Japan, Southeast Asia, Turkey and the Middle East as a teacher, trainer and manager in state schools and higher education for over 20 years. She has a degree in Linguistics, a DELTA, an MA in Linguistics and TESOL, and an MA in Language Testing. She is currently studying for a PhD in Language Testing at CRELLA.

The English component of the University Entrance Exam in Madrid under Review. Questions, Challenges and Future Directions [F]

Miguel Fernández Álvarez

The university entrance examination in Spain is currently undergoing some changes that will be implemented in the near future. Despite the many studies conducted in the past to improve the English component, little changes to the exam have been done in recent

years. A recent study of 3747 students who participated in the EvAU in the Comunidad de Madrid in the academic year 2020/2021 proves that some of the items in the current exam have acceptable item discrimination indices and good reliability coefficients. However, there are other components that need to be revised or removed from the test. The results of the analysis may be helpful in the design and development of the new entrance examination.

Additional Authors: *Jesús García Laborda.*

Miguel Fernández is an Associate Professor in the Department of Linguistics Applied to Science and Technology at the Universidad Politécnica de Madrid. He holds a Ph.D. in English Philology (University of Granada, Spain) and two Masters degrees: M.A. in English Philology (University of Granada, Spain), and M.A. in Language Testing (Lancaster University, UK). His areas of interest include Bilingual Education, English for Specific Purposes and Language Testing.

Lessons from implementing mediation at the European universities based on the MiLLaT project [F]

Sylwia Kossakowska-Pisarek and Malgorzata Swierk

Mediation has been brought to the forefront in language learning and teaching due to the revised version of the CEFR Companion Volume (2020). It is regarded as an undervalued concept and demonstrates the potential for "a reconceptualisation of language pedagogy, curricular reform and the pedagogy of plurilingualism" (Piccardo, 2018: 14).

The paper discusses the key aspects of mediation and the problems encountered during the MiLLaT Erasmus + project (Mediation in Language Learning and Teaching) with the implementation of mediation in the context of Polish, Czech, Lithuanian and Finnish universities. During the project, the teachers prepared tasks for mediation activities and strategies, specified mediation activities and strategies for each task, and then piloted and reviewed them. As the teachers exercised the sole discretion in the choice of tasks, mediation activities and strategies that they prepared and reviewed, we can reach an array of conclusions pertaining to the implementation of mediation. The paper focuses on the impact of mediation on the practical application of language learning and

teaching. Moreover, the challenges are discussed encountered during the implementation of mediation into teaching practice at the higher education level, and implications are explored.

Keywords: mediation, CEFR/CV, language learning and teaching, EU language policy. Council of Europe (2020). Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR): Companion volume with new descriptors. Strasbourg: Council of Europe. <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4> (accessed 27 September 2021). Piccardo, E. (2018). Plurilingualism: Vision, conceptualization, and practices. Handbook of research and practice in heritage language education, 207-225.

Sylwia Kossakowska-Pisarek, PhD, is an assistant professor at the Centre for Foreign Language Teaching, the University of Warsaw, where she teaches Legal and Business English courses. Her interests include mediation, LSP, life-long learning, intercultural communication, e-learning, and positive psychology. She is an author of blended-learning and e-learning courses and some peer-reviewed articles in the area of intercultural communication, e-learning and language assessment. She was also the Research Coordinator for the MiLLaT Project <http://millat.uw.edu.pl/>.

Małgorzata Swierk, PhD. She is an assistant professor at the Center for Foreign Language Teaching, the University of Warsaw. She is the author of papers on e-learning and mediation in EFL. She has been doing her research on the quality of online and blended language courses at HEIs. She is the author of blended, online, and polysynchronous courses. She was awarded the European Language Label in 2009. Also, she was the Coordinator of the MiLLaT Project in 2019-2021.

DELE C1 y C2: prevalencia y puntuaciones de candidatos disléxicos [DI]

Antonia Liberal Trinidad

El Instituto Cervantes, con el objetivo de brindar igualdad de oportunidades para todos los candidatos con indiferencia de la etnia, el género, las condiciones de discapacidad o las dificultades de aprendizaje, adhiere los estándares de calidad de ALTE y SICELE en relación a la adaptación de sus pruebas a los candidatos con necesidades especiales.

En referencia a los candidatos con dificultades de aprendizaje, el Instituto Cervantes recibe anualmente un alto número de solicitudes de adaptaciones de los exámenes DELE para candidatos disléxicos. Estas solicitudes se concentran mayoritariamente en los niveles iniciales e intermedios y provienen, en gran medida, de candidatos en edad escolar. No obstante, también se reciben solicitudes para los niveles C1 y C2, aunque de forma residual. Las preguntas que nos planteamos son: ¿Cómo difiere el perfil de estos candidatos con respecto a los candidatos disléxicos en edad escolar? ¿Muestran las mismas dificultades y en el mismo grado? ¿Qué tipo de adaptaciones requieren/requerirían en función de su perfil y de las características y complejidad de la prueba?

En esta comunicación compartiremos el análisis de los datos de las convocatorias de 2021 y 2022 con el fin de ofrecer una descripción detallada de los candidatos disléxicos que solicitan adaptaciones para los DELE C1 y C2 y de sus necesidades. Para ellos, nos centraremos en el nivel de prevalencia de la dislexia en la muestra analizada y en las puntuaciones obtenidas en las diferentes pruebas, así como su comparación con la del grupo de control.

Antonia Liberal Trinidad has completed an MA in the Training of Teachers of Spanish as a Foreign Language at Universidad de Barcelona and an MA in Language Testing from Lancaster University. Currently, she works in the Certification and Accreditation Department of the Instituto Cervantes (Spain).

Calibration of new rating scales of DELF-DALF: a worldwide collaborative project [F]

Vincent Folny and Marie Gouelleu

In 2005, DELF-DALF diplomas have been aligned to the CEFR and the content of the examinations has been fully refit. However, the rating scales were still partly normative, and their reliability not strongly investigated. In that times, DELF-DALF rating scales have been developed by qualified French experts.

During the last years, France éducation international (FEI) has revised its rating scales. All scales have been revised taking in account the Companion volume (2018) as far it was needed. DELF-DALF descriptors have been adapted from CEFR descriptors. New descriptors and rating scales consider the use of rating scales by local raters. 6900 accredited and experienced raters disseminated in 164 countries calibrated around 200 tasks. A multifaced Rasch

calibration has been done using online questionnaires and an “incomplete Design – Connected” (Eckes, 2009). Each scale has been calibrated by hundreds of raters and each task and criteria calibrated by dozens or hundreds of raters. Data has been analysed for reliability purposes. A synthesis offers a view on the quality of measurement (of these new rating scales and new descriptors), on potential threats on validity (construct irrelevance) and on the perception of this new scales by the raters. This analysis has been a way to make sure that the FEI philosophy strongly oriented to the involvement of local experts is adequate. A new validity argument for DELF-DALF (for speaking and writing) has been developed on the ground of validity, reliability, and ethics (Bachman, 2005, Kunnan, 2020).

The content of descriptors, rating scales, findings and a synthesis of all analysis will be presented. It will be of particular interest for researchers or practitioners involved and interested with the use of examinations in a local context or with the development of rating scales for international examinations.

Vincent Folny is in charge of innovation applied to French exams in France éducation international (FEI). He develops new exams, and he is involved in new developments of DELF-DALF and TCF. He is currently working on the development of an automated scoring of essays for the Test de connaissance du Français (TCF). Vincent Folny is actively involved in the definition and implementation of ALTE's quality management. He leads the ALTE Standing committee, conducts regulatory audits in Europe and train auditors.

Marie Gouelleu manages the production of DELF-DALF examinations. She supervises its writing process and validates the contents of the exams (levels A1.1 – C2) based on the CEFR skills scales and in-house specifications. She has been a key actor in writing the CEFR descriptors and implementing new rating scales for the DELF-DALF examinations. She also trains DELF-DALF examiners worldwide and is often invited by global partners to present France Education international rating guidelines and tools in panel discussions.

Defining Alternative Constructs of Multilingual Assessment in Higher Education: The case of English in contact with other languages in mainland U.S. and Puerto Rico [DI]

Eva Rodríguez-Gonzalez

This paper is part of a larger project that provides practitioners within the field of heritage languages contexts in the Americas with examples of challenges faced by these academic communities in the design and implementation of effective assessment practices. The proposed presentation will provide a multidimensional perspective in which different communities of learners are represented. The first part of the presentation seeks to depict and problematize current research in assessment methods and theories explored in Higher Education to assess language learners. These perspectives include different language learner profiles as well as different learning contexts in Higher Education. It also includes varieties of languages used in the classroom and how educators are using these linguistic varieties to assess their students in both programmatic and instructional platforms.

We will provide one example of challenges faced by two different contexts. The first example examines how learner's judgments about their abilities to organize and perform given tasks influence their reported self-efficacies and ultimately inform curricular and assessment practices (Bandura, 1997). Specifically, the project investigates the self-perceived capabilities of Spanish language students enrolled in second language (L2) and Heritage programs in two domains, speaking and writing, throughout a sequence of Beginning and Intermediate courses at a university in the U.S. Southwest. For the last part of the presentation, we will provide suggestions and recommendation for language assessment in multilingual contexts. We will include implications for teacher training in these types of contexts.

In the concluding remarks, will consider the importance and impact of multiple and diverse assessments of heritage languages contexts. We will provide some suggestions based on the implication of multilingualism and teacher training in higher education.

Additional Author: *Rosita L. Rivera.*

Dr. Eva Rodríguez González is an Associate Professor in the Department of Spanish and Portuguese at the University of New

Mexico. She directs the Spanish Second Language Teaching Program at UNM. She also serves as Assistant Dean for Assessment. She has extensive experience in language teaching, researching and designing curriculum for Spanish-English & English-Portuguese learners in different contexts and levels of proficiency. She has published articles in classroom-based language learning and cognitive approaches to adult Spanish language acquisition.

Consideraciones en torno a la Evaluación en el Aula de las Lenguas desde un Marco de Justicia [F]

Edgar Picón Jácome

La evaluación es un tema difícil en el que los docentes requieren formación y apoyo. En el campo de la enseñanza y el aprendizaje de las lenguas el asunto se torna aún más complejo debido al carácter multidimensional del constructo que se evalúa. Preguntas como qué es lengua, cómo definir el nivel de lengua que se espera que los estudiantes logren en un determinado periodo, cuáles son los procedimientos más pertinentes para llevar a cabo el proceso evaluativo de manera válida y viable, cómo liderar un proceso formativo en el aula, deben ser resueltas por docentes y administradores en el día a día. Teniendo en cuenta lo anterior, en esta presentación compartiré fundamentos teóricos generales y experiencias personales, desde el campo de la investigación en el aula y en conexión con mi experiencia de vida, con el objetivo de promover reflexiones sobre el asunto de la justicia en la evaluación (Shohamy, 2001; Suskie, 2002; McNamara & Roever, 2006, Picón-Jácome, 2013).

En un primer momento compartiré el marco que sustenta el principio de justicia en esta presentación: la validez, la ética, la equidad, y la autonomía (Bachman & Palmer, 1996; Benson, 1997; Brown, 2004; Davies, 2017; Freire, 1997; Little, 1995). Seguidamente, presentaré hallazgos de un proyecto de Investigación Acción Participativa que explora cómo una comunidad de práctica de profesores de una universidad pública colombiana influye en la apropiación por parte del profesorado de un sistema evaluativo que busca la consolidación de una cultura de la evaluación como aprendizaje (Picón-Jácome, 2021).

Durante esta parte de la presentación discutiré estos hallazgos en términos de logros y retos para la promoción de prácticas evaluativas justas en el contexto de la enseñanza del inglés como lengua extranjera en contexto universitario.

Edgar Picón Jácome is a FL teacher from Universidad de Antioquia, Colombia, with an MA in TESOL from Greensboro College, US. He is currently pursuing doctoral studies in education at Universidad Nacional de Rosario, Argentina. His interests regarding research are language assessment and testing, learner and teacher autonomy, action research, and narrative inquiry. His professional experience includes teaching EFL in different contexts, teaching LAL courses at the undergraduate and graduate level, and leading classroom-based assessment curricular innovations.

Comparative judgement of L2 writing in learner corpora: does a narrow proficiency range reduce test reliability? [D]

Peter Thwaites

Comparative Judgement (CJ) is a method of assessment in which judges (who may be experts, peers or even novices) decide which of two pieces of student work is "better". These comparisons can be used to automatically rank each piece of work, and from these rankings a grade can be derived. Research from a range of fields has demonstrated that CJ generates reliable, valid, and efficient evaluations of learner outputs (e.g. Jones et al., 2019; Wheadon, 2022).

This presentation reports on early findings from a project exploring the potential for CJ (in tandem with crowdsourcing methods) to contribute to the assessment of L2 texts included in widely used learner corpora that unfortunately often do not come with reliable proficiency measurement.

A preliminary study suggested CJ's reliability for this purpose. However, in focusing on writing samples covering the full spectrum of L2 proficiency (see also Sims et al., 2020), the study provided optimal conditions for such results. A next step is to find out whether CJ continues to be reliable when learner texts are more homogeneous, and therefore more difficult for judges to compare (van Daal et al., 2019). The study reported here therefore involves the comparative judgement of texts from a narrower proficiency range (roughly B1-C1) from the International Corpus of Learner English (Granger et al, 2020). In addition to exploring the reliability of these evaluations, the paper also offers a first attempt at exploring the validity of CJ for L2 assessment by testing which aspects of learner texts (e.g. length, lexical sophistication etc.) best predict CJ-derived evaluations of text quality.

The results contribute both to learner corpus research, where CJ can potentially enrich learner corpora with accurate measures of text proficiency; and to the more general goal of exploring reliable, valid, and efficient alternatives to rubric-based L2 assessment.

Additional Author: *Magali Paquot.*

Peter Thwaites is a postdoctoral research fellow at UC Louvain, where his work focuses on developing pathways for technologies such as crowdsourcing and comparative judgement to contribute to the development L2 assessment. His PhD, from Cardiff University, dealt with the lexical and distributional underpinnings of word association; his work in this area has been published in several journals.

A comparative study of test-takers' perceptions of paper-based and online language examinations [D]

Róbert Márcz

As a result of the COVID-19 pandemic reaching the country in 2020, the legal framework of accredited language examinations in Hungary was modified: online proctored examination systems were granted accreditation by the Hungarian Educational Authority with the aim of ensuring that the same measurement principles are observed. As a result, foreign language exams can now be taken in two modes of delivery: a) paper-based and b) computer-based.

The aim of the research conducted by the experts of the Hungarian Accreditation Board for Foreign Language Examinations in the spring 2022 was to compare the lived experiences of test takers taking language exams in the two different delivery modes. Applying the non-probability sampling method, the language examination centres accredited in Hungary were asked to forward an online questionnaire to their former candidates who had taken the exam in 2020 and 2021. By the set deadline, 2015 questionnaires were completed. The questionnaire consisted of three parts: 1) background information; 2) reasons for choosing that particular delivery mode and preparation method; 3) personal impressions of the different parts of the exam with open-ended questions. Respondents were asked to give answers for Likert-scale items to get quantitative data and elaborate on their lived experiences to provide an emic perspective. Their written answers were analysed using the constant comparative method to identify emerging themes. Of the 2015 respondents, 81% completed paper-based

and 19% took computer-based exam. To justify their choice of the delivery mode, 50% of the candidates mentioned that they were used to paper-based exams, 40% thought they could concentrate better at the exam venue than at home, 38% thought they would take paper-based exams in the future, and 30% stated they did not want to bother with technical facilities. Results of the comparison of the answers provided by the two groups are presented.

Additional Authors: *Krisztina Babos, Szilvia Dömök, Zoltán Lukácsi, Szilvia Gróf, Magdolna Lehmann, Zsuzsanna Soproni.*

Róbert Márcz has a MA degree in English and History at the University of Pécs. Head of the English Department of the ECL International Language Examination Centre, responsible for item writing, test development and examiners' training. Obtained a PhD in the Doctoral Programme in English Applied Linguistics and TEFL/ TESL at the University of Pécs (2019). Title of dissertation: The Washback of Language Exams on Classroom Practice in Hungary in a Complex Dynamic System Framework. Member of the Hungarian Accreditation Board for Foreign Language Examinations.

Validation of a High-Stakes Test: GA IESOL Multiple-Choice Units [F]

Snezana Mitrovic, Emanuela Botta and Giusi Castellana

The purpose of this study is the validation of the GA Classic IESOL examination at CEFR level B2. Listening and Reading in particular, as criterion-referenced achievement tests, with the aim of reaching the multi-trait, multi-method approach of the examination as a whole. The data analysed at the moment pertain to the Listening and Reading units and have been studied to examine the content validity, reliability of the scores and unidimensionality of the construct. Both reading and listening unit consist of three tasks each (22 items) in four different versions (D-E-F-G). For the listening test, four separate exploratory factor analyses were performed and all revealed excellent values of internal consistency. As for construct unidimensionality, excellent values were identified for versions E, F and G and acceptable for version D. The calibration and the factor analysis identified unidimensional models and good internal consistency values for reading versions E, F and G, while, as a result of the exclusion of an item, an approximate fit was identified for D version (SRMR <0.8 and no large residual) and an acceptable internal consistency value. Despite the fairly small sample size and the fact that the values are likely to increase with

a bigger sample, since the analyses are part of an ongoing process of quality assurance, ways of improving the instrument, based on the D version values of individual items, are investigated more closely. In addition, correlation analyses between the scores of the students who completed both tests are positive and significant (0.621 significant at 0.01). Further steps of the study will be the IRT equating procedure of the four versions.

Key words: listening ability, reading ability, language assessment, international examinations, validity and reliability.

Snezana Mitrovic, GA Italy Academic Manager, has been working in assessment for 15 years, designing and running courses and workshops for teachers, and carrying out a variety of projects, as an examination creator as well. Her main areas of interest are the methodology of TEFL and assessment. Her education includes a PhD in assessment from the Sapienza University of Rome. Her main interests have remained the methodology of TEFL and assessment of English as a foreign language.

Emanuela Botta is a school principal now working as an adjunct professor of Experimental Pedagogy at Sapienza University of Rome. She holds a PhD in Social and Developmental Psychology and Educational Research, with a thesis where she developed the first MST test in Italy. She graduated with honours in mathematics and has done research with INVALSI. She is continuing her research in the field of experimental pedagogy, in particular evaluation, school leadership and inclusion. She has published 24 works.

Giusi Castellana is a researcher in Experimental Pedagogy at Roma Tre University Department of Education, (Rome, Italy). Her research interests cover reading comprehension, metacognitive approach, reading assessment, evaluation and teachers training. She is the author of more than 30 papers, two monographs and a questionnaire on reading strategies "Tell me how you read" which, at present, has three versions (primary, lower and upper secondary school and university students) and which has involved about 10,000 students.

Video Shorts

THURSDAY 27th APRIL

Acces to playlist



Implementing inclusive practices in the assessment of language modules in higher education [DI]

Marion Coderch

Proposals for the implementation of inclusive practices in the assessment of language modules in higher education are often the object of debate among language practitioners. While the need for inclusive assessment practices is acknowledged, assessors often express concerns about the balance between this necessity and the risk of unintended consequences, such as grade inflation and academic malpractice.

Since 2018, the Spanish language programme at the Centre for Foreign Language Study in Durham University has been leading changes to ensure that all students are equally enabled to achieve their learning outcomes and to demonstrate their progress, without the need to make ad-hoc adjustments for individual students with particular requirements or in certain circumstances. On the base

of the principles of assessment for learning and assessment as learning, a model has been implemented where holism, flexibility and choice are combined to give all students the opportunity to showcase their learning at the best of their ability. The boundaries between formative and summative assessment have been blurred to present every assessment task as an opportunity to learn. There is a strong emphasis on the dialogue between assessors and students around feedback: these conversations on personalised feedback are instrumental in helping students act on the comments received, so that feeding back is as important as feeding forward.

This short video will summarise the main features of the system, as well as its reception by the students, who appreciated the flexibility and adaptability of the assessment schedule. It will also highlight the main obstacles encountered along the way –namely, the lack of understanding and the resistance to change shown by colleagues in other academic units and in professional services.

Marion Coderch is assistant professor in Spanish at Durham University (United Kingdom), where she coordinates the Spanish language programme offered by the Centre for Foreign Language Study. She has taught Spanish and Catalan language and culture in various public and private institutions, in the UK and in Spain. Her scholarship interests in the field of language teaching include the development of autonomous learning and the integration of assessment for learning in university curricula.

The Impact of Multilingualism on Language Learning: the case of A2 and B2 Students [F]

Olga Vladislavovna Bukach, Elzara Vasilovna Gafiyatova and Marina Ivanovna Sonyshkina

Despite the growing popularity of multilingual language programmes, their effects on students' performance and competency are not yet well understood. Using a multifaceted approach including contrastive analysis of speaking performance and grammar-vocabulary competency we report on the research-in-progress conducted to verify the hypothesis that a multilingual environment contributes to mastering a foreign language proficiency of B2 students and has minor (or no) beneficial effect on A2 students. The data were collected from 1380 students involved into implementation of National Educational Programme "Adamnar" aimed at creating a "multilingual environment" and extending functions of the native (state) – Tatar – language in the Republic of

Tatarstan. "Adamnar" provides courses delivered in three different languages: native, second and foreign (English).

We performed and contrasted results of three proficiency tests of six categories of "Adamnar" participants, students majoring English: A2 monolinguals; B2 (and above) monolinguals; A2 bilinguals; B2 (and above) bilinguals; A2 multilinguals; B2 (and above) multilinguals. To reveal participants' progress, at the end of each of two semesters we ran proficiency tests based on CEFR principles and Cambridge past papers thus assessing their speaking discourse and grammar-vocabulary competencies. The rating scores and discourse analytic measures demonstrate that 72-hour English language courses had a significantly greater effect on performance of B2 multi- and bilinguals than monolingual students, and analysis of A1 and A2 students did not indicate clear differences between the three, i.e. mono-, bi- and multilingual, categories of participants. We plan to conduct iterative proficiency assessment at the end of each semester to monitor students' progress over the next four years. The research findings enhance our understanding of the role that bi- and multilingualism play in language learning in multilingual settings.

In our presentation we also plan to discuss practical relevance of the research results, new challenges and opportunities for teacher training and development programmes.

Olga Vladislavovna Bukach. Candidate in Philology, Cambridge assessment speaking examiner (center RU172, levels A1 – C1). She is currently responsible for organizing and conducting language testing and assessment of students at the Institute of Philology and Intercultural Communication, Kazan Federal University. Her area of expertise and research interests includes language teaching and assessment, as well as stylistics, translation and interpreting.

Elzara Vasilovna Gafiyatova. Head of department, doctor of Philology, associate professor. She is currently the President of Kazan Regional English Language Teachers' Association. She is responsible for coordinating, analyzing, planning and monitoring the international activities of the Institute of Philology and Intercultural Communication, Kazan Federal University, organization and conduct of entrance examinations for foreign entrants. Her research interests include language teaching and assessment, English language and applied linguistics, professional communication, ESL and academic writing. Her publications appeared in a range of international journals.

Marina Ivanovna Sonyshkina. Doctor of Philology, Professor. She is currently Head of the research laboratory "Text Analytics" at Kazan Federal University. Her research interests include text complexity, reading comprehension and the impact of reader

characteristics on comprehension. She is particularly interested in discovering new ways to assess text complexity and readers' ability to understand texts. She collaborated on projects with researchers and organizations internationally and presented her work at many conferences worldwide; her publications appeared in a range of international journals and books.

What's in a name? Examining the construct of name writing in assessing emerging writing of multilingual kindergartners [DI]

Gordon Blaine West and Jeanne Beck

Name writing is one of the first tasks that young learners (YLs) are taught as they develop their emerging writing skills. Previous studies show that while YLs may be able to write their names and know letters in their name, this ability does not correspond with better alphabetic knowledge, grapheme-phoneme knowledge, or spelling ability (Drouin & Harmon, 2009). Nevertheless, name writing is often incorporated into assessments and constructs of emerging literacy (e.g., Puranik et al., 2019).

In this presentation, we argue that name writing should be considered a distinct construct in assessing emerging L2 literacy due to the unique nature of name writing for YLs. We examine English language writing produced by multilingual kindergartners (N = 1,598) from 25 U.S. states during field testing for a new English Learner identification screener, which is now operational. The current study examines a subset of responses from 300 students to two screener tasks: name writing and narrative writing. Samples were selected to include 100 students at three levels: high, mid, and low proficiency, based on scores from the test's writing subsection. All responses were coded for 21 mechanical features (e.g., sound/letter correspondence, letter formation, capitalization). In narrative writing samples, there was a clear distribution of features, for instance, 98% of high-proficiency YLs demonstrate correct spelling, falling to 66% for mid-proficiency YLs, and only 7% of low-proficiency YLs. In name writing, however, 98.3% of all students used correct spelling, including 96% of the low-proficiency students. These results hold across all mechanical features of the name writing task with no statistical differences observed across the three proficiency levels. Our findings reveal that different constructs of writing may be targeted by different tasks. Implications for the design of assessments for emerging literacy skills are discussed.

Gordon Blaine West is a PhD candidate at the University of Wisconsin-Madison in Second Language Acquisition, where he also works with the Assessment Team at WIDA. His work focuses on test development for young learners and analysis of written and spoken language of young language learners. His work has been published in Applied Linguistics Review, Critical Inquiry in Language Studies, Linguistics and Education, and International Journal of Bilingual Education and Bilingualism.

Jeanne Beck is a PhD student at Iowa State University in the Applied Linguistics and Technology (ALT) program. She is a current elementary English Learner (EL) teacher and served as a WIDA doctoral intern in summer 2022. Her work focuses on second language assessment, K-12 ELs, CALL, and Project-Based Learning. Her work has been published in Assessing Writing, CALICO Journal [forthcoming], Language Assessment Quarterly, MIDTESOL Journal, and TESL-EJ.

Adapting the MLAT-ES into Catalan: when young learners' age is not just any other variable [DI]

Maria del Mar Suárez

Young learners' language aptitude is understudied due to a lack of tests covering this period of life. In addition, young learners, in contrast with adults, are still acquiring their L1 and are as well in the process of learning how to read and write. Consequently, a language-dependent aptitude test should be carefully designed for it to be valid, and results should be taken cautiously considering the test takers' age. An additional challenge is found when the young learners to be tested are bilingual, as is the case of the Catalan/Spanish community in Catalonia.

This presentation aims, therefore, at explaining the process to translate and adapt the Modern Language Aptitude Test – Elementary in Spanish (MLAT-ES - Stansfield et al., 2005) into Catalan (MLAT-EC – Suárez, 2010) so as to obtain a valid aptitude measure across grades. The results show that despite the linguistic proximity between Catalan and Spanish, several issues had to be considered, not only linguistic and cultural but also the learners' age range for the test is addressed to. These include the use of unknown or confusing words, the test font type and size as well as the rest of distractors. We conclude that Catalan and Spanish are certainly close Romance languages, but that does not mean that a light translation of the MLAT-ES into Catalan could guarantee a

perfect fitting test in the minoritarian language for several reasons. Nevertheless, the resulting test has proven to tap into the same factors as the original in several contexts and, therefore, is a valid and reliable measure despite the challenges in its adaptation.

M.ª del Mar Suárez is an associate professor at the Universitat de Barcelona. She belongs to the GRAL research group (PI: Carmen Muñoz), for which aptitude, among other cognitive variables, have been among the variables of interest. She also holds a MA in Higher Education Teaching, and is currently the director of the academic journal Didacticae. In her PhD she was able to participate in the validation of the MLAT-ES, a project led by Dr C.W. Stansfield and D. J. Read. Her latest publications include papers on aptitude, teaching innovation and multimodal exposure to English as a Foreign Language.

CSE-based Dynamic Assessment for English Reading in Rural Northwest China [DI]

Huiling Lyu

Dynamic assessment (DA) differs from the conventional assessments in that it aims to reveal not only an individual's matured or fully developed abilities, but also those that are still under development (Poehner, 2008; Poehner & Wang, 2021). While the former range of abilities can be inferred through an individual's independent performance, the latter can be revealed in his responsiveness when his independent performance breaks down and mediation is needed. Secondary students from rural Northwest China are in need of such tailed mediation due to their disadvantages in economic and educational development. The study addressed their English reading difficulties by offering tailored mediation outside the classroom.

First, reading materials were selected from senior high school entrance examination targeting at three types of constructs: identifying details, making inferences and obtaining main idea. Three researchers were invited to make judgments about item construct, and high agreement was achieved (Fleiss Kappa = 0.823).

Then, to calculate item difficulty, 1,300 students' reading responses were collected in matrix sampling and the 2PL Item Response Theory is applied for equation purpose.

Next, in the pilot dynamic assessment sessions, the mediator provides a series of prompts after students complete reading task. Through interactionist approach, students' reading difficulty were identified and mediations were offered.

In the formal experiment, 42 students voluntarily participated in this experiment as control group (n=22) and experimental group (n=20). Both groups participated in the pre-test and post-test as well as the questionnaire survey for their reading strategies. Result shows that experimental group who participated in the 5 sessions of dynamic assessment performed better than the control group who didn't get involved in dynamic assessment in getting the main idea and making inferences. Questionnaire shows the experimental group have higher rates on the mastery of reading strategies. However, students' accuracy and the number of prompts across the five sessions did NOT change linearly, which reflects students' cognition of reading strategies.

Additional Author: Yaru Meng.

Huilong Lyu is currently a graduate student in language assessment.

Embrace the Future of Minority Language Testing: Insights from Zhuang Language Proficiency Test in China [D]

Andy Jiahao Liu

Historically, various English assessments emerged and are well-researched within the language testing and assessment field (Weir & Saville, 2016). Languages other than English assessments have been consequently underrepresented in the accessible literature. In keeping with the recent growth of interest in assessment for multilingualism, this paper showcases the Zhuang Language Proficiency Test administered across China. Zhuang Language Proficiency Test (or Vahcuengh Sawcuengh Suijbingz Gaujsi, VSSG) serves as a pioneer in standardizing minority language learning in China. As the first minority language test, the three-level VSSG is designed to measure test-takers' knowledge of Zhuang language, promote Zhuang culture, and make decisions related to education and employment. Informed by Bachman and Palmer's (2010) Assessment Use Argument framework, I examine and appraise the current version of VSSG. Drawing on content analysis of VSSG-related documents (e.g., test syllabus, question papers), the review suggests that some of the intended purposes of VSSG have been well-achieved but bring limited positive consequences. The study may bear some implications, both theoretical and methodological references, for assessing minority languages within China and worldwide. As the language testing field continues to put English

assessments in focus, it's also my hope that this review, along with other similar reviews (e.g., Zhao & Liu, 2019), can spark more discussion of the language other than English assessments in this multilingual and multicultural era, hence enriching the diversity in language testing.

Andy Jiahao Liu is currently Assistant Instructor in the English Language and Literature Studies Programme at the Department of Languages and Cultures, BNU-HKBU United International College. Before that, he earned his MA degree in English Studies from the University of Macau.

Using Dynamic Assessment of Writing to Promote Technology-Enhanced Learning in Higher Education [D]

Eleni Meletiadou

Responding to students' request for the use of digitally-enhanced formative assessment practices, this project used a dynamic assessment approach that has been developed within the Vygotskian sociocultural theory of learning. It aimed to develop students' professional skills, i.e., digital skills, in blended learning by promoting experiential learning, improve students' writing performance and favourable attitudes towards learning and support their well-being in HEI especially in the post-Covid-19 era. Fifty final year students participated in this project as part of their module. Adopting a process approach to writing, the lecturer/researcher used three rounds of mediation.

- Round 1 (implicit) consisted solely of a scored and highlighted rubric, not identifying the location or nature of the erroneous parts and was provided by randomly chosen peers) via Padlet to ensure anonymity and allow students to experiment with digital platforms.
- Round 2 (relatively explicit) consisted of narrative explanations of problems provided at the end of each participant's report provided by peers and the lecturer via Padlet.
- Round 3 (most explicit) consisted of comment bubbles/specific comments that showed the student the location of the significant problems, explained the issues, and included recommendations for repair. The lecturer provided this form of feedback.

The lecturer also asked students to provide anonymous feedback via Mentimeter during the implementation to explore their perceptions

of the benefits and challenges related to this intervention. This project aimed to foster inclusion of the increasingly diverse student cohorts due to globalisation, develop students' digital, academic, and professional skills and innovate in tertiary education (HEA, 2020b, 2020c). As universities in the Anglophone world attend to operating on a global stage, linguistic diversity in the sector has intensified. New holistic approaches in language policy and assessment need to replace old traditions in a globalized world.

Dr. Eleni Meletiadou is a Senior Lecturer in EDI at London Metropolitan University. She is an enthusiastic peer learning, writing and assessment scholar and an award-winning researcher and adviser with over 20 years of international experience. She is an established academic writer and has presented her work in various international conferences. She is the Chair of the European Association for Educational Assessment (AEA-Europe) Inclusive Assessment SIG and an Expert member of the European Association of Language Testing and Assessment (EALTA).

Towards AI Assisted Essay Marking: Developing and Evaluating Automated Writing Scoring Models for a General English Proficiency Test [D]

Michelle Chen

This study provides an overview of the stages and considerations in developing automated essay scoring models. The writing test under investigation brings some challenges for machine scoring, including many possible score points (0-12) and a diverse test taker population regarding their language, cultural, and educational backgrounds. Our primary goals are to demonstrate the development steps and evaluate the performance of automatic scoring models by comparing the results with human scoring. Two models were trained, one for each writing task type, using techniques drawing from artificial intelligence, natural language processing, and statistics. To train the models, we selected writing samples covering broad proficiency levels from a diverse test-taker group. After each model was trained (approximately n=260 for each) and validated on a small sample (n=50 for each model), they were applied to score a larger sample (total n=271) for us to evaluate their performance.

The study finds that automated scoring showed satisfactory overall agreement with human scoring. The machine scores showed a

central tendency, and, as a result, the agreement levels are lower at the high and low ends of the score distribution. The results are comparable for all test takers and different gender subgroups. Together, the findings of this study show that automated scoring models are a promising solution to mark essays for the general language proficiency test. At the same time, these findings suggest that there is room for the models to be improved and refined with more input. More studies are needed to evaluate the models and their results to ensure that they are valid and fair to all test takers.

Additional Authors: Alex Volkov, Prometric, Katharine Wallace, Prometric and Paul Edelblut, Vantage Labs.

Michelle Chen is an experienced research psychometrician. She holds a Ph.D. in measurement, evaluation, and research methodology from the University of British Columbia. Her work supports test development, evaluates score validity, and promotes a better understanding of measurements and assessments.

Formative Assessment and Student-Centred Learning: a match made in HEaven. [F]

Sebastian Koźbial

This presentation will exhibit the application of Formative Assessment (FA) and, its key component, formative feedback as tools to support student-centred learning (SCL). This action research study was designed to address the challenges Chinese students face when joining the British higher education system, namely a perceived low engagement with self-monitoring and a lack of independence (Chan, 1999) – two essential characteristics of a successful application of FA (Bennett, 2011). Four small groups (42 students in total) of Year 1, 2 and 3 UG students took part in the research over two semesters, as a part of the credit-bearing Nottingham Advantage Award (NAA) programme. The goal of NAA is to ensure that students have more opportunities to develop attributes and skills and engage in activities that will enhance their overall learning experience within the hybrid context of an international branch campus in China.

This presentation will be divided into two main parts. In the first section, a brief overview of FA conceptions will be included, highlighting the problematic notion that the term FA ‘does not yet represent a well-defined set of artefacts or practices’ (Bennett,

2011:1). In the second part, this presentation will focus on the study conducted at the University of Nottingham Ningbo China. This will include a discussion about the design of the study, the results from two sets of surveys and interviews with students, reflection on the participants’ engagement and academic progress, as well as the key outcome from this action research – an alternative approach to curriculum design with FA at its core.

The presentation will conclude with a new conception of formative approaches to teaching and learning (Formative Routines Framework), particularly having in mind various challenges associated with ‘Chinese learners’ (e.g. the impact of collectivist culture of Chinese society on the perception of a ‘good learner’) that are often at odds with student-centred learning (Poole, 2016). As there is also mention of both learner and teacher training, this presentation could be of interest not only to English for Academic Purposes professionals, assessment designers or module conveners responsible for designing curricula, but also to senior management tasked with staff development.

Additional Author: Bogumila Joanna Kozbial.

Sebastian Koźbial works at the University of Nottingham Ningbo China (UNNC) as the university’s Academic Director, and he is responsible for pedagogical development of teaching staff at UNNC. He is a PhD student at the University of Liverpool, UK. He is a BALEAP Senior Fellow and Senior Fellow of Advance HE. He is interested in approaches to formative assessment (theme of his PhD thesis), teaching methodology and using technology to support engagement within teaching in HE.

Inclusive formative assessment practices (IFAP) in Higher Education: Promoting Education for Social Justice [D]

Eleni Meletiadou

According to the literature, there are very few long-term studies which have implemented more than one formative assessment methods in Higher Education (HE) and more to explore their impact on students’ motivation and writing performance (Panadero et al., 2018). Moreover, very few long-term studies, if any, have actually conducted research in HE using a mixed methods approach.

The aim of the current study was to enhance students’ motivation and writing performance and ensure that no student is left behind irrespective of their background. The project developed and piloted the Inclusive Formative Assessment Practices (IFAP) scheme in HE. This study promoted the Education for Social Justice Framework and explored the beneficial impact of inclusive modern educational assessment practices on student outcomes, overall experience, and continuous professional development.

In terms of this year-long project, 200 first-year students were trained and then encouraged to use peer assessment in their first semester and self-assessment in their second semester to improve their writing skills and professional skills, i.e., collaboration, self-regulation and problem-solving. The researcher collected pre- and post-tests (a short report) and explored any improvement in students’ performance by comparing marks. The researcher also explored students’ attitudes during the implementation using anonymous feedback provided by students via Mentimeter and focus group discussions. This project wishes to inform scholarly debate around inclusive assessment practices that can enhance students’ learning and motivation and cater for their diverse needs.

Dr. Eleni Meletiadou is a Senior Lecturer in EDI at London Metropolitan University. She is an enthusiastic peer learning, writing and assessment scholar and an award-winning researcher and adviser with over 20 years of international experience. She is an established academic writer and has presented her work in various international conferences. She is the Chair of the European Association for Educational Assessment (AEA-Europe) Inclusive Assessment SIG and an Expert member of the European Association of Language Testing and Assessment (EALTA).

Modelling information-based academic writing: A domain analysis focusing on the knowledge dimension [D]

Yu Chengyuan

In this digital age of information explosion, it has been increasingly difficult for academic writers to identify information needs, and to locate, access, evaluate, organize, communicate, and use the needed information for writing. While it has been increasingly recognized that academic writing involves the processes of information literacy and is inevitably an information practice (Yu & Zhao, 2021), existing writing assessment paradigms are still pre-dominantly prompt-based (Zhao, 2022) or only integrate the

reading of one or two passages (e.g., Test of English as a Foreign Language) and fail to reflect the information-involved nature of academic writing. To fill such a gap, the present study reports a domain analysis of information-based academic writing which can inform the development of a new generation of authentic writing assessment in the higher education context. This domain analysis specifically focuses on knowledge, one of the most important dimensions of domain analysis (Chapelle et al., 2008; Mislevy et al., 2003).

Four groups of participants were interviewed in a semi-structured way, including twenty-four graduate students, four subject instructors, four writing experts, and six information specialists. As academic writing is a discipline-specific task (Hyland, 2004, 2008; Wingate, 2006), this study selected students and instructors in the discipline of education. The inductive thematic coding of the verbatim transcribed interview data revealed that information-based academic writing involves three types of knowledge: the knowledge of (1) the discipline, (2) information environment, and (3) academic languages. It suggests that such writing is language-intensive, information-driven, and discipline-specific, and highlights knowledge construction as the primary purpose. Discussion considering the nature of academic writing in higher education and implications for assessment development are also presented.

Chengyuan Yu, a PhD candidate at University of Macau, conducts research on language testing and assessment and academic writing. His publications have appeared in Language Testing, Journal of Academic Librarianship, Asia Pacific Education Review, International Journal of Qualitative Studies in Education, and Asian-Pacific Journal of Second and Foreign Language Education.

Diagnosis and Mediation of Argumentative Writing Ability in L2 Concept-Based Debate Instruction [F]

Hua Fu

Argumentation ability is the key element for critical thinking and also essential for the argumentative essay writing. However, the systematic diagnosis and mediation on it has been under-explored.

This study aims to design the mediation moves based on Toulmin's Argument Model so as to diagnose and promote learners' argumentative writing ability. The mediation design also borrows the principle of gradualness (from implicit hints, clues, to explicit

explanations and answers) advocated by dynamic assessment (DA). The conceptual explanations of argumentation follows concept-based instruction (CBI) in debate classes.

Two intact EFL classes in a Chinese university participated the mediation experiment, with experimental group (EG) receiving concept-based debate instruction plus mediation (N=30), and control group (CG) receiving debate instruction only (N=31). Both classes wrote an argumentative essay of comparable level before and after the experiment. They both practiced writing another three argumentative essays in between. For each of the three practice writing, participants in EG received three rounds of mediation arranged from implicit (e.g., overall evaluation of the essays, pointing out the location of the problems) to explicit (e.g., providing the conceptual explanations about argumentation).

After the experiment, face-to-face semi-structured interviews were administrated to five participants randomly chosen from EG to elicit their conceptual understanding of argumentative procedure. The results of independent-sample t-test showed that the EG outperformed the CG in terms of the holistic scores in the posttest essays. The qualitative analysis of students' essays in the EG showed the improvement of the argumentation elements (e.g., using a variety of argumentative elements) and the quality of argumentation (e.g., the adequacy of evidence). The interview results revealed that participants understood and internalized the concept of argumentation and used it in other essay writing. This study provides insights on the integration of DA and CBI in EFL writing process.

Additional Author: Yaru Meng.

Hua Fu is a PhD candidate in language assessment, including Cognitive Diagnostic Assessment and Dynamic Assessment.

Cross-Country Comparison of English Speaking Ability with PROGOS Test [D]

Masuyo Ando and Yukio Tono

PROGOS is an online 20-minute business English test. It measures test-takers' spoken proficiency of production and interaction using the CEFR-J, a localized version of the CEFR in Japan. The test consists of mostly open-ended questions in five parts including interview, read-aloud, presentation, presentation with graphs and role-play. The test result is provided in a form of a detailed feedback sheet which describes the CEFR-J overall evaluation together with

the CEFR levels of the six qualitative aspects of spoken production and interaction, namely, range, coherence, interaction, accuracy, phonology and fluency.

What makes PROGOS innovative is its dual evaluation system. It has the two versions of evaluation; manual and automated. The manual evaluation was developed based on the CEFR Can-Dos and the test-development theories.

In the manual evaluation, the test answers are evaluated by the well-trained human raters and its quality is being monitored on a daily basis. The automated evaluation system was developed based on the human scores by using a voice recognition API and machine learning technologies. The quality is also monitored on a daily basis.

The test can be taken in the browser and in app. The major advantage of PROGOS' 'fit to the Digital Age' is the increased accessibility. By providing both manual and automated versions, we are able to provide convenient, low-cost and quick test-taking opportunities to those who need to assess their English speaking ability in many countries.

Because of its increased accessibility, the number of test takers has been increasing in Japan and in Asian countries since its launch in June 2020.

In this presentation, we will compare and analyze the test results between Japan, the Philippines, Thailand and Cambodia. By combing the use of AI with the increased accessibility and the data volume, we try to provide new perspectives to assess spoken English proficiency for English learners around the world.

Masuyo Ando serves as Chairperson of PROGOS Inc., Tokyo. The company provides English language training programs and PROGOS English speaking test to corporations and universities. She has 25-year experience in promoting English language training/assessment. She received MA in International Relations from University of Chicago, and MBA, from Leonard N. Stern School of Business, New York University. Her area of Interest is EdTech for business English training.

Yukio Tono is a professor in corpus linguistics at Tokyo University of Foreign Studies. He received his PhD from Lancaster University, UK. He also serves as Director of LINGUA, the World Language Centre at TUFS. His research interest includes corpus approaches to language teaching and learning and the use of the CEFR for language resource development. He is also a PI of the CEFR-J project.

Social Justice Forum

Language tests are often portrayed as instruments that open doors to professional and educational opportunities, thereby fostering integration and democratic participation. While it is clearly the case that passing a language test may grant access to education, jobs, and citizenship, it is equally true that not passing such tests denies other candidates these same opportunities. Language tests are tools of selection that function as both door-openers and gatekeepers within a larger societal system. Language tests are never value neutral but represent and reproduce certain ideas of what counts as acceptable and appropriate language forms and language use. **In this forum we wish to examine the values underlying language tests as well as the different roles that language tests play in the societal system and discuss whether and how high-stakes language tests can be a force for the good.**

To that end, this forum provides a venue for scholars from different disciplines to reflect on the use of high-stakes language tests in society. The presentations will be followed by a discussant tying the topics raised together.

THURSDAY 27th APRIL 2023	
Time	Salón Convención (1) (Live streamed) (In-person: Interpretation into Spanish/English provided; Interpretation into ISL provided on request)
14:30-14:40	Welcome & Introduction: Social justice - barriers and opportunities <i>Cecilie Hamnes Carlsen</i>
14:40-15:00	Mandatory integration requirements for foreigners in the EU: what role for human rights? <i>Zvezda Vankova</i>
15:00-15:20	Language assessment in search of social justice: Is it a chimera? <i>Antony John Kunnan</i>
15:20-15:40	Dialogue and engagement between policymakers and language test developers <i>Laura Schildt</i>
15:40-16:00	Citizenship test exemptions around Spanish and Catalan <i>Kamran Khan</i>
16:00-16:30	COFFEE BREAK
16:30-16:50	Plurilingual Profiles and Multilingual Needs of Adult Refugees and Migrants in Greece and Italy <i>Anna Mouti</i>
16:50-17:10	Language testing and language rights <i>Dina Vilcu</i>
17:10-17:20	Discussion <i>Nick Saville</i>
17:20-17:30	Round-off <i>Cecilie Hamnes Carlsen</i>

Social Justice Forum's Abstracts and Bios

Welcome & introduction: Social justice – barriers and opportunities.

Cecilie Hamnes Carlsen

What is social justice, which role does language and language tests play in relation to social justice, and what professional responsibility and concrete possibilities do language testers have to prevent their tests from being used in ways that contribute to discrimination, racialization and exclusion? These important questions are central to the social justice forum and will be briefly discussed as an introduction to the forum, serving as a background for the following invited presentations.

Cecilie Hamnes Carlsen is professor in second language acquisition at the Western Norway University of Applied Sciences. She holds a PhD in language assessment and has worked for 20 years with the development and validation of language tests for adult migrants in Norway. Her research interests are primarily related to the use and misuse of language tests and the consequences of language tests on test takers and society, in the labour market, in higher education and related to residence and citizenship. She is currently involved in the IMPECT project focusing on the consequences of language requirements for residence and citizenship on low-literate adult migrants.

Mandatory integration requirements for foreigners in the EU: what role for human rights?

Zvezda Vankova

This lecture aims to present the human rights standards of the Council of Europe developed through the case law of the European Court of Human Rights and the European Social Charter, and the legal standards of the European Union in the field of integration of immigrants and refugees. A special focus is placed on civic

integration and language requirements, which can affect the enjoyment of family life and long-term residence by migrants and refugees. The lecture concludes with a discussion on the practical challenges when implementing such integration requirements and their consequences in light of the existing human rights standards.

Zvezda Vankova is researcher at the Law Faculty of Lund University and the principal investigator of the project 'Refugee protection or cherry picking? Assessing new admission policies for refugees in Europe' (2023-2026) funded by the Swedish Research Council. She holds a PhD in Law from Maastricht University. Before joining academia, Dr Vankova worked at the Migration Policy Group in Brussels and Open Society Institute in Sofia. The overall focus of her research aims to examine the interface of legal infrastructures pertaining to human mobility and how their enforcement influences the rights and trajectories of people on the move. More specifically, her research interests lie at the intersection of EU law, Public International Law and empirical legal studies with a focus on human rights, migration, integration and complementary pathways for refugees. She has acted as a consultant on migrant and refugee rights matters to the European Parliament, the Council of Europe's Special Representative of the Secretary General on Migration and Refugees, and the UNHCR.

Dialogue and engagement between policymakers and language test developers

Laura Schildt

This talk will explore how language test developers intervene in the policy-making process. Few studies have examined how language test developers and policymakers interact and language testing literature may even cast language test developers as bystanders whose instruments are misused for construct-irrelevant purposes. In discussing how high-stakes language tests can be a force for good it is necessary to ask to what extent test developers can and do impact high-stakes language testing policies in Europe and how they view the use of tests in this context. The current talk will present results from a comparative study which used an

exploratory design and semi-structured interviews with 28 test executives representing 25 exam boards in 20 European countries. The interviews were transcribed and analyzed in NVivo using double coding (Weighted Kappa = .83) with a priori and inductive codes. A simplified horizontal analysis was run to evaluate responses to the research questions by participant and then a detailed vertical, between-case analysis of the data to look for common themes between respondents by research question. Findings indicate that respondents are regularly in contact with policymakers and provide research-based evidence to policymakers which can help improve the validity, reliability, and fairness of language tests. However, the roles that language testers play in the policymaking process vary greatly from unsolicited critique in countries with less participatory democratic structures to participatory decision-making in countries with more open democratic structures.

Laura Anne Schildt is a PhD researcher at Ghent University, Belgium working on a project entitled, 'High-stakes Language Tests, Migration Policy, and Ethical Codes'. Her primary research interests focus on public policy and validation in language testing. Before coming to Belgium, she worked in the area of language teaching and language assessment in Spain.

Citizenship test exemptions around Spanish and Catalan

Kamran Khan

The Spanish language and civic knowledge citizenship has been in place for a relatively short period of time. This talk will examine preliminary findings from the IMPECT which explores the issue of exemptions from such language testing. The data is generated from responses to a questionnaire on language teacher attitudes to citizenship language testing in various countries. The focus of this talk will be in Spain in relation to not only Spanish but other co-official languages. The talk will be of interest to locally based academics, practitioners and teachers.

Kamran Khan is an associate professor and the director of the MOSAIC research group on multilingualism at the University of

Birmingham. He has specialised in citizenship language testing for over 10 years and has worked in the UK, Spain and Denmark. He was a Marie Curie Fellow and has previously worked on a variety of projects related to issues around citizenship, race and security. He is the author of 'Becoming British: Linguistic Trials and Negotiations' (Bloomsbury, 2019).

Language testing and language rights

Dina Vilcu

Language tests can be seen as instruments often used by (political) decision makers to regulate individuals' access to education, work and public services. People who need to prove their linguistic proficiency might come from within the state borders (e.g., from minority communities) or from outside the country. The purpose of "gate keeping" easily comes to mind in relation to language testing for official purposes. This industry roots and grows most intensively and extensively in affluent countries, which support democracy and non-discrimination, including on language grounds. Are language rights, along with the other human rights, really protected in this context? Decision makers can be helped in assessing their part in this by researchers in the field of languages and by test developers, who often reflect on the social dimension of language testing. This presentation advocates for a change of perspective on language testing, which needs to become a gate opener towards languages which are much less tested in present. If language testing is an instrument, then it needs to become an instrument for learning about different cultures, for preserving language diversity and ultimately for contributing to the preservation and promotion of language human rights. This can be done by encouraging the creation and promotion of tests for smaller languages, which will lead to growing motivation for learning new languages, to creation of resources and to a larger use of the languages.

Dina Vilcu teaches and assesses Romanian as a foreign language (RFL) at Babeş-Bolyai University, Cluj-Napoca. She teaches general language and LSAP courses (for social and humanistic sciences and arts), as well as tailor made courses for interpreters and translators. She also teaches courses on LSP and language assessment at MA level. She is interested in the challenges involved in language assessment for less widely tested languages (LWTL), currently chairing ALTE's LWTL SIG. She also writes about the use of language in critical historic moments (like the Revolution in 1989 in Romania) and in movements of protest (like the ones against the alteration of the justice laws in Romania in 2016-2017).

Plurilingual profiles and multilingual needs of adult refugees and migrants in Greece and Italy

Anna Mouti

The issue of language education of adult refugees plays a very important role in the integration of refugee and migrant populations and the recognition of linguistic repertoires and multilingual needs may form the planning dynamics of relevant programs and tailor-made courses. Greece and Italy share a double role both as host and transition countries, as two of the main EU entry points for refugees and migrants since the 2015 refugee crisis. In this study individual differences related to language will be examined in Greece and Italy and plurilingual profiles, linguistic repertoires and multilingual needs of adult refugees and migrants will be assessed and identified, with a particular focus on L2 Greek and L2 Italian. We will examine ways that social justice links to language needs analysis, multilingual assessment, and language education opportunities-barriers for refugees and migrants.

Anna Mouti is an Assistant Professor of Second Language Acquisition and Applied Linguistics, at the School of Italian Language and Literature at Aristotle University of Thessaloniki in Greece. She also cooperates with the MA Programme "Language education for refugees and migrants" of the Hellenic Open University. It was her involvement in the language education of adult refugees and migrants in Greece that led her to the ALTE -LAMI Special Interest Group and she is mainly involved in the LASLLIAM-LAMI research goals. She is an ALTE Individual Expert Member.

LAMI Forum

LANGUAGE TESTERS' ADVOCACY WITHIN THE MIGRATION CONTEXT



Language Assessment for Migration and Integration

FRIDAY 28th APRIL 2023

Time	Salón Convención (1) (Live streamed) (In-person: Interpretation into Spanish/English provided; Interpretation into ISL provided on request)
14:00-14:10	Welcome and Introduction <i>Lorenzo Rocca and Cecilie Hamnes Carlsen</i> <i>Amparo González Ferrer, Spanish Ministry of Inclusion, Social Security and Migration</i>
14:10-14:30	Keynote: The proliferation of border regimes and language testing for migrants <i>Kamran Khan</i>
14:30-15:25	1. Preventing language test misuse <i>Cecilie Hamnes Carlsen, Cathy Taylor and Sabrina Machetti</i>
15:25-15:45	COFFEE BREAK
15:45-16:05	Keynote: Using assessment for successful language learning: issues to address? <i>Neus Figueras</i>
16:05-17:00	2. Advocacy and vulnerable test takers: refugees, low-literate migrants and minors <i>Lorenzo Rocca, Edit Bugge, Marte Nordanger, Sara Karim, Anna Mouti, Paola Masillo, Stefanie Dengler, José Pascoal and Lorena Bellotti</i>
17:00-17:20	Debate <i>Waldemar Martyniuk</i>
17:20-17:30	Closing Remarks <i>Cecilie Hamnes Carlsen and Lorenzo Rocca</i>

LAMI Forum's Abstracts and Bios

Keynote Speaker

The proliferation of border regimes and language testing for migrants

Kamran Kham

Over the last twenty years or so, there has been an increased focus on borders and containment in relation to greater mobility. Quite often this focus has drawn on wider discourses of security and xenophobia to justify border controls and surveillance. Language tests fit within this wider sociopolitical framework.

In this talk, I will outline three particular forms of border. First, I will demonstrate the increased role of citizenship testing as a means of settlement and sign of integration. Second, I will focus on more ad-hoc forms of assessment from post-citizenship tests to interactions with officials, for example. Finally, I will examine the role of extra-territorial forms of testing that take place in other countries prior to the arrival of migrants in their desired location.

Overall, I will argue that it is important to chart the expansion of language testing and assessment in relation to the broader sociopolitical apparatus and to the varying forms of border proliferation taking place.

Kamran Khan is an associate professor and the director of the MOSAIC research group on multilingualism at the University of Birmingham. He has specialised in citizenship language testing for over 10 years and has worked in the UK, Spain and Denmark. He was a Marie Curie Fellow and has previously worked on a variety of projects related to issues around citizenship, race and security. He is the author of 'Becoming British: Linguistic Trials and Negotiations' (Bloomsbury, 2019).

First topic: Preventing language test misuse

Introducing the topic: Taking action to prevent test misuse

Cecile Hamnes Carlsen

It is a growing challenge for professional language testers that the language tests we develop are being used for purposes other than the intended ones, sometimes with detrimental consequences for individuals and society. What responsibility do we language testers have to prevent test misuse, and what can we do when such misuse is detected? This section includes two ALTE LAMI projects focusing on preventing test misuse: MS18 – the development of a new ALTE minimum standard; and the uneven profile project, which stresses uneven language profiles as the natural state for language users and language learners, and differentiated language requirements as an action to hinder discrimination.

Cecilie Hamnes Carlsen is professor in second language acquisition at the Western Norway University of Applied Sciences. She holds a PhD in language assessment, and has worked for 20 years with the development and validation of language tests for adult migrants in Norway. Her research interests are primarily related to the use and consequences of language tests on test takers and society, in the labour market, in higher education, and related to residence and citizenship. She is currently involved in the IMPECT project focusing on the consequences of language requirements for residence and citizenship on low-literate adult migrants.

ALTE MS 18 – A new minimum standard on responsible test use and preventing test misuse

Cathy Taylor

This talk presents the rationale for the development of the new ALTE Minimum Standard (MS18) for responsible test use. We will look at current practices and how tests may sometimes be used for purposes other than those the developers intended; for instance, as gatekeepers to citizenship, residency and educational opportunities. We will explain how MS18 has been based on the ALTE Code of Practice and ILTA Code of Ethics. The second half of the presentation will give an overview of the ALTE audit process and explain how the MS18 fits into the framework. We will explain the consultation process with ALTE members and the present findings from the MS18 pilots. The presentation will close with a summary of the current situation and the next steps in implementing MS18.

Cathy Taylor is a Senior Academic for English Language at Trinity College London who works in the research department at Trinity. She worked as a teacher, teacher trainer, academic manager and examiner in SE Asia, Europe and the UK for more than 20 years before moving into assessment and testing. She has an MA in Language Testing from Lancaster University and has presented and delivered workshops at international language testing conferences such as EALTA and ALTE. Her current research areas are oral communication and assessment (both in-person and online), and rating scale revision and development.

Uneven linguistic profiles. A LAMI project (2022-2023)

Sabrina Machetti

As recommended by the CEFR and the CEFR Companion Volume, to best support second language learners it is necessary to give careful consideration to their individual competence profiles.

The presentation focuses on the development of a LAMI project aimed at supporting institutions and policy makers in making informed decisions in the fields of language assessment policy and uneven linguistic profiles. The paper presents an overview of key issues, concepts and processes related to the research area under consideration, with reference to the European recommendations.

Sabrina Machetti is Associate Professor of Educational Linguistics at the University for Foreigners of Siena (Italy). Her research interests focus on L2 Italian testing and assessment, and language policy in education, specifically in relations to migrants. At the same University, she is the Director of CILS Centre (Certification of Italian as a Foreign Language).

Keynote Speaker

Using assessment for successful language learning: issues to address?

Neus Figueras

Assessment specialists have worked for decades on the development of learner-relevant assessment approaches that can contribute to successful language learning. It seems, however, that the potential of classroom assessment has not yet been fully taken advantage of.

This session will briefly address the reasons for this continuing challenge in the field and discuss the issues to address in order to make learning-oriented language assessment work. The issues facing the construction of a possible road map for using assessment for successful language learning entail untying the Gordian knot binding formative and summative assessments, and can be grouped into three main strands: conceptual, methodological, and practical. Each need to take into consideration the role of different contexts and different stakeholders, as they play an important part in the choice of available options and in how they are implemented.

Neus Figueras has been involved in a number of international research and development projects related to assessment at different education levels and has given courses and presented in universities in Spain and in various other European countries, as well

as Asia and the USA. She has also carried out consultancy work in assessment and in curriculum design. She has collaborated regularly with the Council of Europe in the uses and the dissemination of the CEFR in language testing and assessment and is one of the authors of the Manual for Relating Language Examinations to the CEFR (Council of Europe, 2009). She has recently co-edited with David Little Reflecting on the CEFR and its Companion Volume, published in 2022 by Multilingual Matters, and contributed to the publication of Aligning Language Education with the CEFR: A Handbook, jointly published by the British Council, UKALTA, EALTA and ALTE, publicly available online. She is a founding member of EALTA (European Association for Language Testing and Assessment, www.ealta.eu.org), and is currently the coordinator of the EALTA CEFR SIG. She is a member of the organizing committee of GIELE (Grupo de interés en evaluación de lenguas en España, <http://giele.webs.upv.es/>).

Second topic: Advocacy and vulnerable test takers: refugees, low-literate migrants and minors

Introducing the topic: Teaching & assessment according to a broader view: human rights and quality education

Lorenzo Rocca

The contribution addresses the Council of Europe (CoE) vision of Europe “as a multinational and multicultural society, where immigrants take part as equal members, on the basis of equality of rights and opportunities in return for equality of obligations” (CoE, 2003). According to this vision, high-quality education is not only an important means to reach these goals, but a fundamental right in itself; a right to be preserved for all, including – or better, starting from – the more vulnerable migrants, such as minors, refugees, low literates.

Professionals can play an important role, at different levels and with different interlocutors, on the one hand providing evidence related to constraints to the access to education, on the other hand contributing to tailoring tools to the specific needs of vulnerable migrants. Among these professionals, language testers are asked for constant engagement, e.g. in raising awareness with the stakeholders, improving connections between teaching and assessment, and developing assessing materials for the learning environments.

Based on this perspective, the talk will point out the need for language testers to commit to more than the guarantee of test quality. Examples will be given to support this declared need, referring to the work of LAMI in the LIAM activities and introducing the aims of the IMPECT project.

Lorenzo Rocca is responsible for the language projects department at the Società Dante Alighieri in Rome. Engaged in test development and teacher training for over 20 years, his research interest is focused on the migration context; in light of this focus Lorenzo has been the chair of the ALTE LAMI SIG since 2008 and a member of the Council of Europe LIAM project since 2014. In 2020 he was awarded with the national scientific qualification Associate Professor.

Migration tests and low-literate migrants: preliminary outcomes from the IMPECT project

Edit Bugge, Sara Karim and Marte Nordanger

Norway introduced language and knowledge requirements for citizenship in 2017, and dating from 2022, the language requirement was raised from A2 to B1 for oral skills. While the political and ideological underpinnings of citizenship requirements have attracted considerable attention in research, less is known about their real-world consequences for test-takers. A key objective of the IMPECT project is to provide insight into the lived experience of test-takers, and to what extent the prospect of passing the tests and becoming citizens motivates language learning and integration, or if such requirements may hinder development and ultimately lead to exclusion. This presentation, drawing on two independent interview studies, puts the voices of low-literate adult learners of Norwegian center stage: one consisting of interviews with learners from various country backgrounds who have completed the Introductory Program and face the requirements of the process

of becoming citizens (Bugge & Nordanger), and the other focusing on Kurdish migrants' experiences with migration tests, their sense of belonging, and investment in language (Karim). Preliminary outcomes imply that requirements may motivate language learning on some level, but also create a persistent space of liminality and produce feelings of uncertainty, discouragement and distrust that inhibit and, sometimes prematurely, disrupt learning.

Edit Bugge leads the IMPECT project. She has a PhD in sociolinguistics on intergenerational language transmission, and her research interests include language policies and social conditions for language variation, and adult second language and literacy development in social contexts.

Sara Karim is a PhD Candidate in the IMPECT project at Western Norway University of Applied Sciences. She is a sociologist by training, and her research interests include migration, identity, belonging, and second language learning of adult migrants.

Marte Nordanger is currently a post doc researcher in the IMPECT-project. She has a PhD in second language acquisition, and her research interests include adult second language development, the development of literacy, and the interplay between social and cognitive aspects of learning.

LLAT, the ALTE-LAMI practical tools for teaching and assessment based on The Council of Europe LASLLIAM Reference guide

Anna Mouti, Paola Masillo, Stefanie Dengler, José Pascoal and Lorena Bellotti

This presentation focuses on the presentation of LLAT, the four tools representing the last multilingual deliverable of the LAMI SIG.

The talk will start by describing the broader frame in which such tools are embedded – the piloting of the Council of Europe LASLLIAM reference guide – and their main aim, to illustrate the use of LASLLIAM for assessment purposes within the learning environment, specifically regarding the beginning of the learning process (the welcome phase) and the end-of-course exam.

The contributors will speak on behalf of the 18 LAMI members, from 12 ALTE institutions. They will share the floor in order to present in

a nutshell the process that led to these new instruments targeted to low-literate migrants and designed to support the concept of profiles and the “recognition of partial competences” (CEFR, 2001: 175). A short description of the contents of each tool will be followed by some data related to the qualitative validation phase. Strengths and limitations will be pointed out, as well as the future perspectives of the LLAT project.

Anna Mouti is an Assistant Professor of Second Language Acquisition and Applied Linguistics, at the School of Italian Language and Literature at Aristotle University of Thessaloniki in Greece. She also cooperates with the MA Programme “Language education for refugees and migrants” of the Hellenic Open University. It was her involvement in the language education of adult refugees and migrants in Greece that led her to the ALTE -LAMI Special Interest Group. She is an ALTE Individual Expert Member.

Paola Masillo has a PhD in linguistics and teaching Italian as second language from the University for Foreigners of Siena. Her main research focuses on learning Italian as a foreign/second language, language assessment, and language policies. She has been a post-doctoral research fellow at the CILS Centre of the University for Foreigners of Siena, where she is currently working in the technical-scientific and data processing area.

Stefanie Dengler is an employee of the Goethe-Institut, the Federal Republic of Germany's cultural institute. She works in the department of examinations and is responsible for test development. In ALTE she is an elected member of the Board of Trustees and the Standing Committee.

José Pascoal joined ALTE in 1991 as the University of Lisbon representative. He has been involved in many ALTE projects and publications, namely within the ALTE LAMI and YL SIGs. He co-founded CAPLE-ULisboa, where he designed and developed the CEFR six-level examination system, its adaptation to schools, the A1-B1 diagnostic test TEJO for 9-11 pupils, and the A2 oral examination for illiterate migrants applying for Portuguese citizenship. He designed and wrote the content for the Online Portuguese Platform (High Commission for Migrations). Currently based at the University of Macau, he has provided consultancy to international language organizations as well as to the Portuguese and the Macau Education Departments.

Lorena Bellotti is Vice-President of KCE (the Kultura Centro Esperantista), and represents in ALTE the Esperanto culture and language. She has dedicated her whole life to the world of languages, by attaining a degree in languages, and building huge experience as second language teacher and language tester.

Debate

How to foster appropriate test use in the areas of access to education/schooling/higher education, access to the labor market, as requirements for residency and citizenship

Moderator: Waldemar Martyniuk

Waldemar Martyniuk is Professor 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 Visiting Professor and lecturer at several universities in Germany, Switzerland (Basel), and in the USA (Stanford University). He was also seconded to the Council of Europe, Language Policy Division in 2005-2006, and in 2008-2013 as Executive Director of the European Centre for Modern Languages in Graz, Austria, and is Chair of the Board of Trustees at the Association of Language Testers in Europe (ALTE).

ALTE would like to thank all Sponsor Organisations who have supported this conference:

