aea '22 EUROPE

New Visions for Assessment in Uncertain Times

09-12 November, 2022, Dublin, Ireland

Book of Abstracts

Book of Abstracts

1: Unlocking 'Assessment Cultures'

9:00 - 16:30

Unlocking Assessment Cultures

S. Shaw[1], E. Andressen[2]

[1]University of Cambridge, United Kingdom
[2]Andressen Byram Ltd, United Kingdom

What exactly is an ‘assessment culture’? And what insights can we gain into our own assessment practices by comparing them with practices and cultural influences in other countries and contexts? Through a series of thematically-linked contributions led by members of the AEA-E Assessment Cultures SIG Steering Group, this workshop will provide a stimulating forum for participants to reflect upon these issues. The workshop will explore the influences and impacts of the cultures in which we live and work on the ways in which assessment is and has been conceptualised, developed and used in different parts of the world, as well as in different disciplines, organisations and contexts. The workshop will also afford an opportunity for presenters and participants to draw together both specialist and experiential knowledge about this emerging field and to contribute to the academic development of what it means in practice to describe the ideas, customs, attitudes and behaviours which make up an assessment culture. Through highly interactive sessions, the workshop will seek to determine how far our own assessment practices have been influenced by and adopted both across and within international assessment practices.

Put your test to the test: Assessing test quality

9:00 - 9:55

Put your test to the test

C. Sluijter[1, 2], B. Hemker[1]

[1]Cito, Netherlands
[2]Fontys University of Applied Sciences, Netherlands

Educational tests serve a specific goal, such as evaluation, monitoring, diagnostics, selection or guidance. Such a goal is only met, if the test is of sufficient quality. This workshop aims to provide participants with practical tools to evaluate the quality of a test.
Our target audience consists of people involved in test development. Participants should have experience with at least some of the elements of test production. They also should have an understanding of the basic psychometric principles of testing.

In the theoretical part of the workshop we give an overview of evaluation systems and show their similarities and differences.

In the practical part of the workshop we put the theory to practice, by having participants actually evaluate the quality of a test of their own choice, based on relevant information pertaining to the test. This includes possible research reports on how the norms are determined, and the reliability of test scores and the validity of test interpretation and use. The workshop leaders assist participants in applying the evaluation criteria to their own test.
In the final discussion, the findings of each participants are discussed and we round off with a list of practical lessons learned.

Item Banking and the Assembly of Test Forms

9:00 - 16:30

Item Banking and Assembly of Test Forms

A. Verschoor[1], S. de Klerk[1]

[1]Cito, Netherlands

The workshop will offer an introduction into Item Banking and applications for test assembly from a practical point of view. Participants will gain insight in the do’s and don’ts when using an item bank for the purpose of developing assessment instruments, and will receive practical guidelines to use metadata and psychometric theory to assemble optimal tests based on an existing item bank. participants will have hands-on experience in using automated tools to make linear or adaptive tests, based on Item Response Theory (IRT) or Classical Test Theory (CTT). Main features of these applications will be addressed in the workshop. Participants will be able to understand and assess the usefulness of item banking in their own work.

Thursday,
10 Nov, 2022

Keynote Speech

10:45 - 11:30

The Impact of Test Items Incorporating Multimedia Stimuli on the Performance and Attentional Behaviour of Test-Takers

P. Lehane[1]

[1]Dublin City University, Ireland

Technology-Based Assessments (TBAs) use items that employ a broad array of interactive, dynamic or static stimuli e.g. simulations, animations, text-image. Although it is assumed that these features can make TBAs more authentic and effective, their impact on test-taker performance and behaviour has yet to be fully clarified.

This research investigated the extent to which the use of different multimedia stimuli can affect test-taker performance and behaviour using a mixed methods approach. Guided by four main research questions, an experiment was conducted with 251 Irish post-primary students using an animated and text-image version of the same TBA of scientific literacy. Eye movement and interview data were also collected from subsets of these students (n=32 and n=12 respectively) to determine how differing multimedia stimuli can affect test-taker attentional behaviour. A second study involving 24 test-takers completing a series of simulation-type items was also undertaken. Eye movement, interview and test-score data were gathered to provide insight into test-taker engagement with these items.

The results indicated that, overall, there was no significant difference in test-taker performance when identical items used animated or text-image stimuli. However, items with dynamic stimuli often had higher discrimination indices indicating that these items were better at distinguishing between those with high and low levels of knowledge. Eye movement data also revealed that dynamic item stimuli encouraged longer average fixation durations on the response area of an item. An examination of the data relating to test-taker performance and behaviour for simulation-type items found that there was a weak to moderate relationship between task performance and time-to-first-fixation on relevant information/areas.

Education systems around the world are now attempting to devise their own TBAs for their terminal post-primary exams e.g. New Zealand, Ireland. It is hoped that the findings of this research will act as a resource for those who wish to use TBAs in this manner. In particular, insights into test-takers' eye movements may help to support more appropriate inferences from test scores.

Poster Session

11:30 - 11:45

Evaluating the Impact of Self-Assessment as an Assessment for Sustainable Development Strategy in Higher Education

E. Meletiadou[1]

[1]London Metropolitan University, United Kingdom

The higher education literature testifies to an extensive interest in self-assessment (SA) because it emphasizes learner responsibility and metacognitive skills. SA also aligns perfectly with the aims of Assessment for Sustainable Development (ASD) as it meets the needs of the present without compromising the ability of students to meet their own future learning needs. Moreover, SA promotes learner-centered pedagogy which sees students as autonomous learners and emphasizes the active development of knowledge rather than its mere transfer and/or passive learning experiences. The present case study, which used a mixed-method approach, explored the use of SA as an ASD practice aiming to help students improve their writing performance, self-regulation, and motivation. In terms of this semi-experimental study, forty-four undergraduate students used SA of writing for one semester. Students' pre-test and post-test scores showed that SA improved undergraduate students' writing performance significantly. The outcomes from students' pre- and post-implementation surveys revealed that SA had a strong impact on students' self-regulation. Finally, considering the findings from students' focus group discussions and self-reflective reports, the current study concludes that SA was challenging but it considerably improved students' critical thinking and sense of personal accountability.

A learner centred approach to digital assessment item type design and development

S. Mistry[1], S. Mistry[1]

[1]Cambridge Assessment International Education, United Kingdom

Cambridge International conduct numerous user research trials as part of the digital assessment research and development strategy. The poster will illustrate the proven methodology being used, how it is essential for success and emphasises the importance and value of a learner centred approach, observational evaluations, and iterative design protocols. There are definitive steps of the trialling process, from desktop research, material validation, school engagement through to design thinking, user testing and iterative changes based on outcomes. Throughout, it is essential to understand what success looks like. Experience shows the factors and key issues need considering when using international contexts for both face to face and remote user testing approaches.

Examples highlight important trials that underpin the discussed approach, including a significant piece of ongoing research investigating how young learners (age 6-12) interact with different on-screen item types, in terms of cognitive and motor skills, in turn informing design and functionality considerations. Findings to date have lay the foundations for subsequent item prototyping with learners and additional studies into on-screen item type design and functionality. This is across a range of subjects and includes trials of subject specific item types such as those used in Computing.

11:30 - 11:45

The decision making processes of examiners of performance assessments

C. Scully[1]

[1]Dublin City University, Ireland

The assessment of practical skills is a key component of numerous undergraduate programs. However, the rapid switch to online assessment brought about by COVID-19 posed distinct problems regarding the testing of practical skills. When assessments are designed and administered, it is crucial that evidence is documented so that decisions made on the basis of scores are deemed valid. As such, the move to the online testing of practical skills necessitates ongoing research into the validity of these assessments.

One area requiring attention is the decision-making processes that examiners of such assessments go through when determining a student's competency level. This is particularly important when the assessment is delivered online for high-stakes purposes – uncovering the "how" of examiner decision-making should strengthen the validity argument.

This poster will report on an ongoing mixed-methods study regarding the assessment of nursing Objective Structured Clinical Examinations (OSCEs). A series of students were filmed completing two OSCEs: blood pressure measurement and naso-gastric tube insertion. These videos were shown to 12 assessors, who vocalised their thought processes and completed marking guides. The resulting qualitative data will be used to determine how assessors make judgements, while the marking guides will be used to calculate reliability between assessors.

Understanding the demands that digital tests make on teachers' assessment literacy

G.A. Nortvedt[1], K.B. Bratting[2], H.H. Haram[1], O. Kovpanets[3], A. Pettersen[3]

[1]University of Oslo, Norway
[2]Universitetet i Oslo, Norway
[3]UiO, Norway

Teachers' assessment responsibilities and tasks change when they move from the paper-based format to digital configurations. For instance, providing instructions to students and scoring and grading their assessments, formerly the teacher's responsibilities, are tasks often left to the test delivery platform for digital assessments. Traditionally, these tasks provided teachers with insights into the challenges their students encounter, as well as their capabilities, which the teachers used to support their interpretation of the assessment data. However, teachers may have more difficulty interpreting assessment data from digital tests and, thus, find using assessment data formatively to provide student feedback and plan teaching interventions more challenging. This poster presents data from teacher interviews, focusing on the teachers' experiences with a national level formative numeracy assessment and their reflections on interpreting data from a digital assessment compared to their previous work with paper-based assessments. The purpose of the study was to explore how transitioning from teacher-administrated paper-based assessments to digital assessments changes what teachers need to understand about assessments and about how to use assessment data. Potential consequences for professional development will be discussed.

“Assessing students' non-cognitive skills: Nazarbayev Intellectual Schools approach"

A. Rakhimbekova[1], Z. Rakhymbayeva[2]

[1]Nazarbayev Intellectual School, Kazakhstan
[2]Nazarbayev Intellectual Schools, Kazakhstan

The labor market is changing. The work force of the future will need “people skills”—meaning they will need the sophisticated emotional, interpersonal, and cognitive skills necessary for working in teams, managing complex tasks, sifting, and sharing information, and collaborating. In recent years, scholars, practitioners, and the lay public have grown increasingly interested in measuring and changing attributes other than cognitive ability. These so-called “non-cognitive” qualities are diverse and collectively facilitate goal-directed effort (e.g., grit, self-control, growth mindset), healthy social relationships (e.g., gratitude, emotional intelligence, social belonging), and sound judgment and decision making (e.g., curiosity, open-mindedness). Longitudinal research has confirmed such qualities powerfully predict academic, economic, social, psychological, and physical well-being.
The aim of this research is to identify and assess Nazarbayev Intellectual Schools (NIS) students «non-cognitive» skills. NIS is an experimental platform for the development, testing, implementation, monitoring of modern models of educational programs for secondary education levels in Kazakhstan. Based on the international experience in the formation of a framework for assessing «non-cognitive» skills of students, we have developed a framework for assessing those skills of NIS students. We assume that the assessment of non-cognitive skills implies an analysis of the teachers’

11:30 - 11:45

A Digital Number–Line Estimation Task: Scoring and Implications

H.H. Haram[1], K.B. Bratting[2], O. Kovpanets[3], G.A. Nortvedt[1], A. Pettersen[3]

[1]University of Oslo, Norway
[2]Universitetet i Oslo, Norway
[3]UiO, Norway

In 2022, the third generation of the Norwegian mapping tests in numeracy will be published and used by most 1st and 3rd graders in Norway. The tests aim to measure students' early mathematics development and are intended to identify students at risk of falling behind. One of the items developed for the mapping tests is the number–line estimation (NLE). NLE tasks are often used as part of number sense tests, and performance on NLE tasks has been found to correlate highly with other more advanced mathematical competence measures. We investigated whether NLE tasks can differentiate between students who draw on their proportional reasoning when estimating and those who use less advanced strategies, such as counting-based strategies. This poster will show how process data was used to analyze students' answers to set dichotomous scoring thresholds that could possibly differentiate between students who used different strategies to perform estimations. Furthermore, our poster will discuss how practical constraints in the digital test platform and the need for transparency and simpler interpretations for the teachers have shaped the development of the NLE tasks.

11:30 - 11:45

Language assessment practices in the COVID-19 era in Greece

D. Tsagari[1], T. Liontou[2], C. Giannikas[3]

[1]Oslo Metropolitan University, Norway

[2]Department of English Language & Literature/ National & Kapodistrian University of Athens, Greece

[3]Cyprus University of Technology, Cyprus

The abrupt global outbreak of the pandemic in early 2020 impacted the majority of the language education community and led to Emergency Remote Teaching (ERT) in all fields of education. The rapid planning and implementation raised issues and questions among practitioners and researchers worldwide especially in terms of student assessment. As part of an international survey on educators' perceptions and online teaching and assessment practices during the pandemic, this presentation focuses on English language teachers' assessment practices within the Greek educational context. It examines issues pertinent to teachers' assessment literacy skills through the lens of complex and unpredictable adaptive systems. More specifically, this presentation will discuss 300 EFL educators' attitudes towards summative and formative methods of assessment during the pandemic and their perceived feelings of satisfaction and self-efficacy when facing the challenge of accommodating and modifying their language assessment techniques to cater for the rapidly changing contextual characteristics during the COVID-19 pandemic. Our survey findings are expected to shed light on the emotional and pedagogical challenges the Greek ELT community faced in order to re-imagine their assessment practices in digital learning environments and share the lessons learned.

11:30 - 11:45

Annotation consistency, measured: A methodological poster

F. Morley[1]

[1]Cambridge University Press and Assessment, United Kingdom

With the theme for this year's conference being "New Visions for Assessment in Uncertain times", it is worth revisiting a traditional area of marking practice, but viewing it through a modern lens. Annotations exist as a cognitive tool for markers and an accountability tool for supervisors, reviewers and centres. However, no one has attempted to measure annotations with a statistical metric which can measure annotation consistency efficiently and broadly at different levels of analysis, without having to arduously look through individual papers. To do this the study draws from the literature in reliability and explains how this can be applied to annotation consistency. Each measure has costs and benefits, both in terms of validity and interpretability. On top of this, the results of the measure must be communicated with appropriate visualisations which represent the metric both accurately and clearly. This study aims to develop this metric through discussing different methods to measure and visualise annotation consistency, and how to compare this across different groups of interest like subjects, markers, and item sizes.

Comparative judgment as a formative assessment activity in legal education

E. Hartell[1], K. Egelandsdal[2]

[1]KTH Royal Institute of Technology, Sweden
[2]Department of Education, University of Bergen, Norway

The poster will present a research project on the use of adaptive comparative judgment in legal education with the purpose of supporting law students' understanding of quality work. Comparative judgment in this context is about letting students compare different administrative decisions in pairs. Approximately 300 law students will be involved during an administrative law course. Groups of 30 students will assess administrative decisions of varying quality. During the iterative process of pairwise comparison of decisions; students chose which one is better, and justify their choice based on assessment criteria for legal method, language, and content. Using learning analytics, the judgment of these legal decisions will be ranked for each student group in terms of quality. This ranking along with the student comments and experiences from the assessment activity will serve as a foundation for group discussion about quality in legal decisions in public administration. The project aims to investigate whether and how comparative judgement may contribute to support the students in developing a better understanding of variations and levels of quality in legal work. Data will be collected through observation, semi-structured group interviews, learning analytics and a student survey.

11:30 - 11:45

Validation of large-scale high-stakes tests for college admissions decisions

P.J.(. Ho[1]

[1]University of Oxford, United Kingdom

It seems inevitable that the higher the stakes, the more demanding the validity requirements. In the case of large-scale tests for college admissions decisions, however, evidence of validation is rarely brought to public scrutiny. There is therefore a need to investigate how validation of college admissions tests has been carried out, as well as how the validation efforts have been communicated. This research consists of three primarily qualitative studies. The first study is a systematic literature review that explores varied approaches to test validation in high-stakes settings. Preliminary literature searches support anticipated differences among the approaches to validation, which parallels the absence of unanimity over the conception of validity. The second study considers the Hong Kong Diploma of Secondary Education Examination as a case study, featuring an independent validation of the regional assessment of English language proficiency by constructing and evaluating validity arguments. It would empirically unearth the congruities and dissonances of the intended and actual score interpretations of the test. The third study continues to be situated in Hong Kong, using critical discourse analysis to explore the notion of transparency in the context of large-scale high-stakes testing and the extent to which such demands are met by policy discourse.

Irish primary school teachers’ mindset and approaches to classroom assessment

J. Malone[1, 2]

[1]University of Oxford, United Kingdom
[2]Rethink Ireland, Ireland

Systemic assessment reform requires consideration of, and engagement with, key stakeholder perspectives. This new research examines Irish primary school teachers’ implicit beliefs about learning, or what Dweck (2006) calls ‘mindset’, and their approaches to classroom assessment. Building on research by DeLuca, Coombs and LaPointe-McEwan (2019), this 2022 study explores how teachers’ mindsets might influence their approaches to classroom assessment similar to the way that mindset has been shown to influence teachers’ pedagogical practices. The working hypothesis is that teachers with a growth mindset, who believe intelligence and talent are dynamic, will tend towards formative and student-centred approaches to assessment, whereas teachers who believe that intelligence and talent are fixed will tend towards more summative and standardized approaches to assessment. Data is being collected during the period March-May 2022 through an online survey, which combines the Theories of Intelligence Inventory (Dweck, 2008) and the Approaches to Classroom Assessment Inventory (DeLuca, Valiquette, Coombs & LaPointe-McEwan, 2018). This research presents insights into Irish primary school teachers’ current perspectives on learning and classroom assessment as the effects of the pandemic are still being felt in Irish classrooms, and provides stimulus for discussion about the future of teacher education and professional development programmes.

11:30 - 11:45

Developing and validating a model of whole-class violin teaching in primary schools.

M. Déiseach[1]

[1]Trinity College Dublin, Ireland

The in-school practice of whole class learning of a musical instrument is currently increasing. While this is a positive development, it is being conducted on an ad-hoc basis. Tracking of children’s technical progress is rare (Murphy et al, 2011).
This research is aimed at developing a measurable approach to the in-school whole class teaching and learning of one instrument; the violin. Commonalities of technical approach to violin skill acquisition have been identified across a variety of established violin methods. Three principle challenges of Whole Class Violin Teaching (WCVT) have been collated; classroom management, the performance strand taught in isolation, no cohesive learning goals (Fautley et al, 2019).

The use of Key Performance Indicators in the music classroom have long been established as an effective vehicle for formative assessment. Research has shed light on a number of problems with these indicators in the design process; identifying and isolating sequential indicators of performance, mapping indicators to the domain, building consistency in the assessment process. A set of KPI specific to WCVT is currently being drafted. The design problems with KPI in the general music classroom should predict the challenges of developing this assessment tool for specific WCVT.

11:30 - 11:45

Assessment policy in education for England and Scotland 1998-2018.

M. Taylor[1, 2]

[1]University of Glasgow, United Kingdom

[2]Liverpool Hope University, United Kingdom

In 1998 events occurred which started an unprecedented volume of research and policy publications on the topic of assessment in England and Scotland. The publication of Assessment and Classroom Learning (Black and Wiliam, 1998), and the devolution for Scottish powers in the Scotland act (1998) all followed the election of the New Labour Government in 1997.

This research poster details my critical discourse analysis of the key policies between 1998 and 2018, concerning the descriptions of how assessment is purposed in the two nations of the United Kingdom during times of political change. This work explores the patterns of change in how assessment might be deployed as a tool for accountability and national accounting in the name of raising standards.

In reflecting on these twenty years of change, we can see how various education policymakers have tried to define what is important in education and often measure that through different iterations of assessment processes. In looking ahead for assessments in the future, it is important to consider the past changes so we can learn relevant lessons and create directions to enable assessments processes to be more in line with the learning that we are trying to measure.

Implementing a national assessment system in Angola

M. Borges[1], A. Lobo[1], A. Monteiro[1], M. Gomes[1], J. Veloso[2]

[1]Institute for Educational Assessment, Portugal

[2]Instituto Nacional de Avaliação e de Desenvolvimento da Educação, Angola

“National Exams Pilot – Angola 2022” is a bilateral cooperation project between the Ministry of Education of Angola and the Ministry of Education of Portugal. Coordinated by IAVE – Institute for Educational Assessment (Portugal), the pilot in Mathematics and Portuguese Language, set forth in 2022, covers a sample of 2100 students, from all provinces, attending grades 6 and 12. This project aims at creating the grounds for the implementation of a national exam system in Angola after 40 years without standardized assessments. Based upon the growing trust placed by the Angolan government in assessment results for evidence-based decision making, the project involves setting up a sustainable strategy for the generalization of national standardized assessment, in order to perceive the status of curriculum development in the primary and secondary school levels.

The project includes the definition of strategic guidelines as well as an action plan with activities for each key process. There are four areas of intervention: (i) definition of an organizational model for national assessments; (ii) instructions for conducting national exams and other guidelines; (iii) human resources training in item and test design, marking, database design, statistical analysis of results and reporting; (iv) printing, distribution and test application processes.

11:30 - 11:45

Language Assessment against the backdrop of the COVID-19 Pandemic

K. Vogt[1], D. Tsagari[2]

[1]University of Education Heidelberg, Germany
[2]Oslo Metropolitan University, Norway

The recent Covid-19 pandemic resulted in major disruptions, including school shutdowns which have triggered a shift in educational instruction as teaching and learning move online, creating a variety of challenges for educators, esp. in terms of assessment provision and accommodations. A recent report published by UNESCO (2020) showed that assessment of students' performance has faced challenges. Teachers, in particular, had to meet an imperative to embrace pandemic pedagogies, digital literacy, and alternative approaches to assessment. This state of affairs has challenged conceptualizations of assessment and conventional test delivery modes.

In this poster, we discuss the possibilities and affordances created due to the demands of emergency remote assessments and explore the potential for meaningful professional development of teachers' assessment perceptions and practices. A study was undertaken with higher education teachers within a research community of language assessment combined with questionnaires. Critical reflections and group discussions combined with prior responses to an open-ended questionnaire assisted teachers in thinking about the effectiveness of their assessment practices, ways of improving them and promoting their language assessment literacy. Through this study, we shed light on teachers' experiences with Emergency Remote Assessment realities which can be instrumental for professional development in terms of language assessment literacy.

11:30 - 11:45

Assessment as a pedagogical tool: Wellbeing in the wake of the pandemic

I. Suto[1], C. McKenna[1], H. North[1], C. Jellis[1]

[1]Cambridge Centre for Evaluation and Monitoring, United Kingdom

Background:
Mental wellbeing is an important part of personal and social education. The pandemic has increased global awareness of it, and of the need for teachers and children to understand it. We converted a wellbeing research instrument into a pedagogical assessment for classroom use.
Theory:
McLellan and Steward (2015) define wellbeing as feeling and functioning well, uniting two psychological theories: of hedonic and eudaimonic wellbeing. They developed a research instrument and used it with 40 English schools (5170 students).
Methods:
We extended the instrument into a 22-item digital assessment with improved coverage of hedonic and eudaimonic wellbeing. We created linked age-appropriate lesson plans. In each plan, teachers taught Part 1, administered the assessment, then taught Part 2. We trialled the pedagogical process with 1200 children aged 7-16 across 18 diverse schools, obtaining feedback through a teacher questionnaire and interviews.
Findings and discussion:
Many teachers appreciated the assessment's pedagogical value. Others used it to evaluate class wellbeing, to take a holistic approach to understanding children, and to identify individuals with low wellbeing. We argue these multiple purposes are complementary.
Reference:

McLellan, R. & Steward, S. (2015) Measuring children and young people's wellbeing in the school context, Cambridge Journal of Education, 45:3,307-332.

11:30 - 11:45

Contextual and psychological correlates of national exams during COVID-19 pandemic

N. Curkovic[1], L. Lukačin[1]

[1]National Centre for External Evaluation of Education, Croatia

State matura, as standardized national exam, is one of the main methods of selection for majority of universities in Croatia. The aim of this research was to determine how the changes caused by COVID-19 pandemic related to results of State matura exams in Mathematics and Croatian language. Sample of 9307 high school graduates filled out the online questionnaire in which their motivation for State matura exams, psychological well-being and learning experiences during COVID-19 pandemic were assessed. Learning experiences were assessed by subscales of the PISA Global Crises Module (GCM) that focuses on describing changes in education practices caused by pandemic and included students' activities, used resources, feelings related to learning from home and problems with self-directed learning. Motivation was assessed with EVC (Expectancy-Value-Cost) scale and well-being was assessed with CORE-YP scale that measures psychological distress. The results of the study indicate low, however still important role of motivation and psychological well-being for results of high-risk examinations. Moreover, the results show negative impact of pandemic on various aspect of learning experiences.

11:30 - 11:45

A Policy Document Analysis of Post-Soviet Assessment Policy in Kazakhstan

R. Kakabayeva[1]

[1]King's College London, United Kingdom

In 2015, a gradual transition to a new criteria-based assessment policy consisting of formative and summative assessment models began in mainstream schools in Kazakhstan. From the first day of its introduction, the new assessment policy has been a controversial and much disputed subject within society. This study provides an overview and analysis of the assessment policy documents underpinning criteria-based assessment policy. The study outlines educational assessment policy in Kazakhstan through the analysis of a wide range of documents that have contributed to the history, development, implementation, and evaluation of the new assessment policy with the aim of exploring the links between these and the policy imperatives framing new visions of assessment.

Document analysis revealed a very detailed image of social, economic, political, and ideological shifts as well as the origins and growth of new assessment visions, including their aims, methods, and desired outcomes, as well as their applicability to school life. Similar thoughts were extracted from a document analysis conducted by newly created quasi educational officials, which emphasised the creation of favourable conditions in schools for the implementation of new evaluation methods. Following that, document analysis allowed for a deeper dive into the many viewpoints on the new assessment system.

11:30 - 11:45

An instrumental approach on students’ work with digital items in mathematics

M. Winnberg[1]

[1]Stockholm University, Sweden

This poster proposal presents an early-stage research study investigating students work on innovative spreadsheet items from an instrumental approach. This theoretical lens scrutinizes the interrelations between technology and mathematical thinking. The overall research aim is to shed light on important factors that need to be taken into account whenever similar items in mathematics are applied in summative digital assessment, in this case the Swedish national tests in mathematics for Grade 9. The purpose of this study is to examine students’ instrumented techniques, along with instrumented action schemes in the process of digital genesis. The research is framed by technical problems, students’ described opportunities and constraints, and also how this relates to their conceptual understanding.

One of the arguments of digital assessment in mathematics is construct broadening. At the same time it is crucial that tests assess mathematical competencies, and do not introduce construct-irrelevant variance. In this context the tests need to emphasize the intended mathematics and not computer skills. The importance of studying students work with digital items is thus based on validity concerns. The presented study pinpoints such concerns related to instrumental techniques applied (or not applied) by students when solving items with a spreadsheet component.

11:30 - 11:45

Modeling Extreme Response Styles using IRTrees in attitudinal scales from TIMSS 2019

A. Christiansen[1], R. Janssen[2]

[1]IEA Hamburg, Germany
[2]KU Leuven, Belgium

Self-reported rating scales are widely used in background questionnaires of (international) large-scale assessments to measure constructs that may explain achievement differences. However, when answering such questionnaires, participants may adhere to an extreme response style (ERS) in order to boost their responses (Böckenholt, 2013). The presence of ERS introduces systematic errors in the measurement of attitudinal scales and therefore reduces construct validity (Khorramdel, Davier, & Pokropek, 2019; N. Kim & Bolt, 2020). The present poster describes an IRTree approach (Jeon & De Boeck, 2019) to investigate such extreme response styles. These multidimensional models allow to disentangle the primary trait of interest from possible ERS by assuming that respondents choose their response to an item on a rating scale through a sequential, multi-stage cognitive process. The framework is applied to the self- reporting surveys of motivation towards science and mathematics as applied to 8th-grade students during the 2019 cycle of the Trends in International Mathematics and Science Study (TIMSS). It is shown that there are systematic differences among students in their choice of extreme response categories but this variable is not related to students' background characteristics nor to their test performance.

Change in country rankings in PISA after filtering out examinees who engage in rapid guessing

M. Michaelides[1], M. Ivanova[1]

[1]University of Cyprus, Cyprus

Test scores may be valid indicators of individual achievement or aptitude, assuming the test-takers were motivated enough and engaged with the test content. With international assessment programs which are low-stakes for examinees, it is likely that some will not put adequate effort when responding to the test. Thus, their scores underestimate their achievement, and aggregate country scores are biased downwards. The current study examined how country rankings could be impacted after filtering out examinees who engaged in rapid guessing in the 2015 computerized PISA Reading assessment. Rapid guessing at the item level was identified with a normative approach, i.e., responses occurring in less than 15% of the mean item response time in the country sample, and alternatively with a fixed 5-second threshold approach. Filtering out examinees who rapid guessed at 5% or more of the test items, country means slightly increased, but the median change in country rankings was only 1.75 (normative method) or 1 (fixed method) position. Of the 56 countries, only 10 (normative) or 6 (fixed) had a larger than 5 positions change in their rankings. Overall, the impact of removing rapid guessers on country rankings was small.

11:30 - 11:45

Perceived difficulties in mathematics in relation to test performance

A. Engström[1]

[1]Umeå universitet, Sweden

Learning difficulties in mathematics (LDM) is a contested research area. Educational and non-educational sciences represent different theoretical paradigms resulting in divergent views concerning the origin and nature of the difficulties. Consequently, procedures in educational settings tend to vary. In assessment situations, especially when the stakes are high for the examinees, this could interfere with fairness principles. This paper proposes a theoretical model, investigating and conceptualising different aspects of LDM in relation to high-stakes assessment. The suggested model draws on the multidisciplinary feature of the field and is operationalised through a self-report instrument concerning perceived difficulties in mathematics. The participants are test-takers of the college admission test, the Swedish Scholastic Aptitude Test (SweSAT). The analyses focus on the instrument itself and its relation to test performance. In addition, the study investigates the eventual effect on the constitution of the test taker group due to COVID-19, focusing the representation of low achievers on the SweSAT.

The aim is to provide additional knowledge about LDM and to offer a basis for further studies, scrutinising fairness and validity issues in high-stakes assessment practices, e.g. LDM-related accommodations. Findings support a three-dimensional model, and a relation between self-reported math difficulties and SweSAT test-performance.

11:30 - 11:45

Does the level of in-person school attendance during the COVID-19 pandemic explain differences in grade 6 math performance in Flanders? Evidence from a national assessment and administrative data.

S. Spikic[1], M. Goos[2], R. Janssen[1]

[1]KU Leuven, Belgium
[2]KULeuven, Belgium

Math performance among Flemish primary school students has been declining for some time, leaving the region vulnerable to the COVID-19 pandemic disruption of the education system. Aside from nationwide school closings, students were forced to stay home due to their class or entire school being quarantined. Also, during certain periods physical presence in class had to be limited to a certain number of students, depending on the size of the classroom. Consequently, the level of in-person school attendance during the pandemic varies between students, classes and schools. This study examines if differences in math performance among students from the final year of primary education can be explained by the amount of distance education students were faced with. Unique data is used that contains the 2021 national assessment of math performance, as well as administrative data on school closings from the Flemish ministry of education.

11:30 - 11:45

Are 21st century skills teachable and assessable?

S. Miller[1], Y. Bimpeh[1]

[1]AQA, United Kingdom

21st century skills have become one of the popular topics in education. Along with a growing emphasis on including them in educational programmes, there is a need to be able to assess students’ competency in these skills via high-stakes exams. While there is no single generally accepted definition of 21st century skills, Salas-Pilco describes them as the skills 'that empower learners to enable them to cope with the demands of the present century'.

These 21st century skills are not novel; they have been variously known over the years as transferable skills, employability skills or essential skills. It was not until recently, however, that policy makers and educators recommended distinctly including them in academic content. Notwithstanding the general consensus about the relevance of these skills, challenges continue to exist with regard to the definition, description, assessment and teaching of these so-called 21st century skills.

In this discovery stage of our research, we summarise academic literature in order to explore how teachers teach 21st century skills and how we assess these skills. As we outline the concerns of researchers and practitioners, we emphasise the urgent need to move beyond conceptual frameworks toward pragmatic solutions.

11:30 - 11:45

Supporting learning with remote formative assessment

D. West[1]

[1]AQA, United Kingdom

This presentation explores whether we can apply the same subject and assessment expertise that underpins GCSE and A-level qualifications to strengthen support for students throughout their learning journey. Due to events over the past two years, learners are no longer confident that they will progress in a linear path towards high-stakes summative assessments.

AQA is exploring how changing the nature of the feedback provided to students can support different approaches to assessment. We also consider if, and how, we might place formative pedagogy into a digital environment.

Classroom learning is often encouraged by oral questioning and feedback discussion, but this can be supported remotely by tailored resources to assist learning tasks, assess knowledge, skills and understanding, and provide guidance to correct misconceptions. When a student starts a new topic, there is scope to provide immediate feedback on their incremental learning outcomes, either via a remote platform or a teacher in the classroom. As the complexity of a task increases, speed of response may be less important, as detailed feedback related to the individual student’s work will engage their attention. The essential factor is that any feedback should be understood by the learner so that they can respond to it effectively.

11:30 - 11:45

Are they still learning? - what happened when the classroom became a screen

P. Simoes[1], J. Cachucho[1]

[1]IAVE, Portugal

Exceptional conditions caused by the COVID-19 pandemic, especially the lock down and suspension of face-to-face teaching in the second half of the 2019/2020 school year, brought about the need to monitor and assess the different conditions in which the students and teachers might have worked and learned, and also the quality of the learning itself and the eventual drawbacks of distance teaching and learning. In the same way, it was considered of extreme importance the characterization of the context in which the teaching and the learning took place as well as the individual circumstances with which the students had to cope during the lockdown.

Given the contextual background, a diagnostic study of the quality of the ISCED 1 and 2 students' learning was commissioned to IAVE (the Portuguese office for educational and large-scale assessment) by the Portuguese Ministry of Education. Considering the time allocated to the development of the study and the time of the school year it would be carried out, it was decided it would be fully computer-based and administered to a sample of students.

A Study on the Identifiability of the Logistic Positive Exponent Model

J. Gonzalez[1, 2]

[1]Pontificia Universidad Catolica de Chile, Chile

[2]Millennium Nucleus on Intergenerational Mobility: From Modelling to Policy (MOVI), Chile

The logistic positive exponent (LPE) model is an item response theory model that extends the regular 1, 2 and 3PL models by incorporating a parameter accounting for asymmetry of the item response functions. Although the model has been used in applications, particularly using a Bayesian approach for the estimation, the identifiability of the model parameters has not yet been established. In this paper we conduct an identifiability analysis of the fixed-effects version of the LPE model. The results show that the model is not always identifiable. Thus, although the LPE model has been rediscovered and used in practice in assessment, this paper shows that technical requirements, such as model identifiability, are needed to ensure correct and useful inferences that lead to good policy practices.

Psychometrics and Test Development I

13:45 - 14:15

How to choose the anchor test when equating test scores

I. Laukaityte[1], M. Wiberg[2]

[1]Umeå university, Sweden
[2]Umeå University, Sweden

Test score equating is used to make scores from different test forms comparable. A commonly used data collection design is the nonequivalent group with anchor test (NEAT) design, where we have two samples of test takers from two nonequivalent populations who each take different test forms and a common anchor test form. The anchor test score has a crucial role as it is used to examine differences in ability level between the groups. Although some research has been conducted on how the anchor test should be comprised there are still some uncertainties. The overall aim was to examine how different anchor tests affect the equating transformation using both real data and simulated data. We used different anchor tests given to several administrations of the Swedish Scholastic Aptitude Test (SweSAT). The SweSAT is a multiple-choice binary scored test which is used for college admissions and is typically given twice a year. In addition to real data, we used simulations based on the SweSAT administrations to examine different conditions. The preliminary results show that the equated values vary depending on the used anchor test. Practical implications as well as recommendations for how to choose anchor tests for standardized achievement tests are given.

14:15 - 14:45

The quality of students' educational attainment can't always be quantified

A. Scharaschkin[1, 2]

[1]University of Oxford, United Kingdom
[2]AQA, United Kingdom

Many approaches to educational assessment assume that educational attainment is a phenomenon that can be validly represented as a quantity. It is assumed that variation in the quality of students’ performances, with respect to some assessment criteria, is appropriately represented as variation in quantity on a numerical, equal-interval scale.

To explore empirically the question of whether attainment has quantitative structure, methods of conjoint measurement theory were applied to dichotomous test data from national examinations in physics and economics in England in 2018.
It was found that for these assessments the assumption that the data could be used to create an equal-interval scale was highly unlikely to hold.

This presentation will argue that a more general conception of measurement, that does not require location on a numerical continuum, is more appropriate for the assessment of constructs of this kind.

An extension of the conjoint measurement methodology to appraise the results of assessments consisting of polytomous items (far more common in national testing in England) has been developed and will be discussed.

The question of classifying educational attainment into ordinal classes without assuming an underlying quantitative structure will be briefly discussed .

Data-driven direct consensus standard setting without IRT

M. Van Onna[1], C. Jongkamp[1]

[1]Cito, Netherlands

In general, the use of data enhances the quality of standard settings, as experts get realistic feedback on their estimates of item or test difficulty for borderline students. Keuning, Straat & Feskens (2017) proposed a data-driven direct consensus setting method (3DC), where experts have to indicate cut scores on clusters of items. The experts get visual feedback on the appropriateness of the alignment of their cut scores with the actual difficulty of the clusters by means of a figure. Keuning, Straat & Feskens (2017) derived these estimates from a fitted one-parameter logistic model. In this presentation, a more classical approach is presented. The relative difficulty of cluster scores is based on the observed score distribution. This approach results in similar estimates as the 3DC-method when only one exam form is administered, and a Rasch model fits the data. In addition, this new approach can also be used when the data are not suited for IRT-modeling, as we will demonstrate with an application of the new method in standard settings of Cyprus teacher admission exams.

Education Policy & Assessment I

13:45 - 14:15

A long weekend in Summer 2020 - exams in crisis

T. Oates[1]

[1]Cambridge University Press & Assessment, United Kingdom

After Pandemic struck the globe in 2019-20, exams were cancelled in England and a model established to award grades in ‘general academic qualifications’ (GCSE and A Levels) to individual 16 and 18 year-olds. Developed by the national exams regulator in conjunction with exam boards and with the oversight of the Department for Education, the model was used to calculate grades over a four month period prior to results release day in the second week of July 2020. The managed public process was thrown off course by limitations of the model and an announcement by Scottish politicians that they were abandoning a similar model. With Scotland’s processes running in advance of the process in England, this exerted considerable political pressure on decision-makers there. The events are an important case study in the management of exams in crisis - which can originate from causes other that Pandemic, such as natural disaster, social and political upheaval, conflict or administrative breakdown. Such seismic breakdown may occur again, and rather than seeing the Pandemic as ‘once in a lifetime’, assessment professionals may benefit from understanding the detail of events, in order to understand how to develop robust contingent arrangements.

Developing and implementing MERV – a conceptual model for investigating, envisaging and evaluating the balance of different, but complementary principles of assessment within qualifications.

P. Johnson[1], C. Taylor[1]

[1]Qualifications Wales, United Kingdom

As the national Regulator with responsibility for qualifications and the qualifications system in Wales, Qualifications Wales has an important role in conceptualising and articulating different models of assessment within qualifications. Research undertaken as part of our qualification reform and development work has identified a deficit in considerations around manageability and engagement from the perspective of educational institutions and learners. This has prompted further research into how assessment can be more engaging for learners and whether specific forms of assessment are manageable for teachers.

The findings of this research have enabled us to develop MERV - a conceptual model for considering the most appropriate balance of Manageability, Engagement, Reliability and Validity in qualifications in Wales. The original idea for MERV has evolved from an abstract concept into a usable paradigm that encapsulates different aspects of our qualification reform and development work. We have now taken this a step further and tried to operationalise our use of MERV by creating a toolkit for developing a new qualification. This toolkit uses MERV as the overarching framework for testing and evaluating the probable impact of different aspects of a qualification’s overall design and has the potential for wider application and use.

14:45 - 15:15

Assessment, policymakers and communicative spaces - striving for impact at the research–policy interface

L. Gray[1], J. Gracey[1]

[1]AQA, United Kingdom

To achieve a new vision for assessment, educational assessment experts and researchers must actively engage with policymakers, education stakeholders and the public, ensuring the right conditions are in place to give new policies and practices the greatest chance of flourishing. Despite the lessons of the pandemic, reform in educational assessment systems has proved difficult to achieve, especially around high-stakes examinations.

This presentation will examine how exam board assessment researchers can interact with the policymaking context in which they work. We will present a model of successful research–policy influence, drawing on the Habermasian concept of ‘communicative spaces’. This theory advocates that to achieve influence on educational assessment policies, the exam board assessment researcher has to plan, carry out and communicate their research, not to policymakers, but with policymakers. We consider not only the question of what we communicate, but who communicates it, how, and to whom. These issues go to the heart of our conceptualisation of exam board assessment research as a communicative space.

The presentation will conclude by sharing some examples of these practices in action, and will show how one educational organisation has sought to achieve uptake of its research and achieve impact with policymakers and wider educational stakeholders.

Summative Assessment

13:45 - 14:15

Estimation of marks for technical qualifications using different methods

Z. Rahman[1]

[1]City & Guilds, United Kingdom

Estimation of missing marks for an assessment due to learner absence has been more widely debated since the start of the Covid-19 pandemic, following the cancellation of most assessments in the UK in 2020 and 2021. During these uncertain times, where Coronavirus continues to impact learner absence, mark estimation is even more important in enabling continuation or completion of study.

A research study was undertaken to investigate the accuracy of the z-scores, percentiles and proportional mark estimation methods for several Technical Qualifications. This enabled the review of correlation between estimated/actual marks as well as the percentage of learners getting the same estimated/actual grade. Many similarities were seen between the different methods but differences were observed for different scenarios and qualifications. In general, it was observed that using both the spring exam and assignment more reliably estimated marks for the summer exam, and that the assignment and the spring exam provided a less accurate estimate of each other.

This paper presents the findings from this research, encourages a discussion on what may be considered an acceptable level of accuracy, and aims to contribute to the limited research in this area for vocational assessments.

The effects of using testing and restudy as test preparation strategies on educational tests

Z. Hao[1], J. Baird[1], Y. El Masri[1]

[1]University of Oxford, United Kingdom

When students self-prepare for a test, restudy and testing are two commonly used test preparation strategies. We investigated the effects of testing and restudy strategies on test performance in authentic test preparation contexts. We conducted 3 experiments in which students reviewed educational psychology concepts by either restudy or testing and took a post-intervention test immediately, followed by a final course test five days later. Across the three experiments, we manipulated test preparation strategy use (testing vs. restudy), the stakes of the test (low-stakes vs. high-stakes), and metacognitive training (yes vs. no). We found that when students prepared for a high-stakes test, both testing plus feedback and restudy produced better performance in the post-test than did the control (no preparation) group. Testing plus feedback also outperformed the control group on the delayed final test. The testing and restudy, however, did not appear to have substantial advantages over the control group on low-stakes test performance. Students in the metacognitive training groups received psychoeducation, made monitoring judgments, and managed their study attempts in addition to restudying and testing. As indicated by greater post-test and final course test scores, the restudy effect was amplified by metacognitive training during test preparation.

14:45 - 15:15

Assessment in non-standardised oral exams – examiners´ arguments for, negotiation of and legitimation of decisions in grading

M.S. Syverud[1], T.S. Prøitz[1]

[1]University of South-Eastern Norway (USN), Norway

In Norway, oral exams are non-standardised summative assessments, and grades students receive from these exams are included in their final diplomas. Since admission for higher education is based on these diplomas, oral exams are high stakes for students. The oral exams consist of a prepared student presentation and a conversation between two examiners and the student. Examiners must then decide an agreed upon grade for the student. However, little is known about how examiners argue for, negotiate, and legitimise their common decision regarding student performance. Applying thematic analysis on video-recordings from 36 grading conversations between examiners in two lower and two upper secondary schools, this paper presents a qualitative study of how examiners assess students in oral exams. The research questions are: How do teachers assess students in oral examinations? What arguments are used, and how are decisions legitimised? Preliminary findings indicate that examiners start the decision-making process with vague and general comments on students' performance, before posing more concrete arguments both linking variably to pre-defined curriculum goals and concerning students handling of the exam situation. The study revisits issues raised in previous research on what counts in summative assessment, and what examiner roles teachers take during summative assessment.

International Assessments

13:45 - 14:15

Open Book Exams: what do we (really) know about their impact on learning and assessment?

R. Hamer[1], J. Jacovidis[2]

[1]International Baccalaureate, Netherlands
[2]Inflexion, United States

In 2020, The International Baccalaureate commissioned Inflexion to carry out a review of the existing evidence on effects of and good practices for assessment using reference aids (a.k.a. open book exams, OBE). The review showed that literature is fractured, and that empirical research has been unsystematic and focused on single classroom or institute cases. Studies differ greatly in focus, design, data and analysis. As a result, the empirical evidence is divided on the positive impact of OBE on learning, student well-being and assessment, with outcomes depending greatly on the way it is implemented.

While the literature review resulted in a set of guidelines for schools and organisations to consider when implementing successful OBE, it offered very little information on schools and their needs regarding preparing students for OBE, especially in an international context. A survey of staff, educators and schools showed significant support for increasing the OBE options on offer for summative assessment. Authors will present the main findings of the review and the survey as well as an overview of good practice guidelines for OBE, and introduce of a suite of multi-year OBE pilots in international schools informing any future OBE offerings.

14:15 - 14:45

The impact on the performance of 15-year-olds in Ireland on the PISA reading, mathematics, and science tests when testing occurs at two different periods in the same year (spring vs autumn)

S. Denner[1, 2], G. Shiel[3], M. O'Leary[2]

[1]Educational Research Centre, Ireland
[2]Dublin City University, Ireland
[3]Educational Research Centre, Dublin, Ireland

The Programme for International Student Assessment (PISA) assesses the performance of 15-year-old students. In Ireland, the administration of PISA currently occurs in the spring which is a very busy period in Post-Primary schools. In moving PISA administration to the autumn it is important to measure if there is an impact on the performance in reading, mathematics, and science. To measure the impact, data were collected in spring 2018 (Main Study) and in autumn 2018 from another sample of schools (Feasibility study). Data were also collected from a sample of principals and a sample of English, mathematics, and science teachers who participated in the autumn study to gain insights into any factors that might impact the move.

Results from multi-level modeling indicate that, while controlling for the school and student background variables and other level-1 variables that are associated with performance and which were examined in this study, ‘time of year’ when PISA testing occurs does not impact on overall student performance. However, significant interactions between ‘time of testing’ and ‘gender’ on mathematics performance were observed while controlling for student- and school-level variables. In understanding these results, insights from teachers and principals are examined.

14:45 - 15:15

Gender gap in mathematics in France: comparing results from TIMSS advanced and the Baccalauréat.

F. Salles[1], M. Le Cam[2]

[1]Ministry of Education DEPP, France
[2]Conseil d’évaluation de l’école, France

In France, the DEPP implements international surveys. Alongside these specific assessments, certificative national examinations are milestones in students’ schooling. These two types of assessment are different in terms of methods and representation students have of them. It makes sense to compare domain specific students’ performance across these different types of assessments. How does students’ performance measured by both types of assessment correlate? Does this correlation vary depending on students’ characteristics? In the mathematical domain in particular, surveys regularly inform us about the gender gap in mathematics. Does this gender gap depend on the assessment characteristics? Does it vary depending on whether the assessment is high or low-stakes for the student?

The DEPP conducted a study to compare the performance in mathematics as measured by two different assessment instruments: the international TIMSS Advanced and the baccalaureate. The TIMSS Advanced 2015 plausible values were used with the baccalaureate scores of these same students. Correlation analyses and regression models were applied, controlling for demographic variables. The results of this study show that the gender gap in mathematical performance is about six times larger on TIMSS Advanced than on the baccalaureate mathematics test. They suggest that the context of assessment influence performance.

Assessment Against the Backdrop of Covid I

13:45 - 14:15

The value to teachers of digital, classroom-based assessments in a post pandemic world – A case study.

S. Mistry[1]

[1]Cambridge Assessment International Education, United Kingdom

A study was conducted to understand the value to teachers of digital progression tests and associated reporting. The study focussed on digital topic-based tests, across IGCSE Biology, Chemistry and Physics. Using outcomes from an initial small-scale trial from 2019, schools were given free access to the tests for use with learners as required, across an 8-month period during 2021. Feedback from surveys, a forum and focus groups, highlighted how teachers used the tests to support teaching and learning activities, together with aspects of test design, delivery, and reporting. International learners, from 30 schools took over 9000 tests.
Teachers saw value in the tests for:
- Evaluating pupil learning and understanding at both beginning and end of a topic,
- Use as a revision/ preparation tool for mock or final exams,
- Use as a ‘formative’ assessment to establish students readiness for moving to the next unit/ topic,
- Highlighting common misconceptions in understanding,
- Creating prompts for further classroom discussion,
- Use as a back-to-school resource when returning from home learning,
- The detailed reporting and ability to give written feedback.
Further research will take the outcomes to further develop and improve digital assessment offers in this assessment space.

14:15 - 14:45

Is the math performance decline in Flanders steady for different kinds of students? A closer look from the Grade 6 math national assessments

M. Goos[1], S. Spicik[1], J. Denis[1], R. Janssen[2]

[1]KULeuven, Belgium
[2]KU Leuven, Belgium

This study examines time trends in the math performance of Grade 6 students in Flanders, tackling the questions whether covid-19 has strengthened the decline in math performance among Flemish primary school students, and if this is the case, among which subgroups of students the decline was the steepest. Data from 3 cycles of national Grade 6 math assessments (2009, 2016 and 2021) were used for this purpose, and analyzed via a series of two-level logistic regression analyses. Results seem to show steeper declines in particular math domains and among specific groups of students. Venues for improvement of math class practice will be discussed.

E-Assessment I

13:45 - 14:15

Progressive achievement approach in formative assessments

S. Krstic[1], H. Claydon[2]

[1]ACER, United Kingdom
[2]ACER UK, United Kingdom

Using examples of one of our assessments, ACER will focus on presenting why the progressive achievement approach in assessments is important. These types of assessments that are based on progress achievement approach show each pupil their learning progress. This is accomplished through three steps. Firstly, data is used as evidence to locate where the pupils are in their learning. Secondly, the information from these assessments can inform teaching and learning through enabling teachers to set pupils' objectives, as well as focus their resources and their own development. Lastly, data from these types of assessments can be used to track each pupil's progress, by measuring growth in pupil's achievement over time. These steps are operationalised through the assessment reports, all of which contain descriptive proficiency scales, which describe the nature of growth in the area being measured. The description focuses on the skills, knowledge and understanding that pupils typically demonstrate at this point. These reports illustrate how learning progresses, so that teachers can make evidence based decisions on planning their lessons to best suit their pupils' needs. Lastly, these kinds of reports are not only useful for teachers, but for pupils too as it may help them own their own learning.

14:15 - 14:45

Standard setting across multi-mode qualifications

L. Miller[1], S. Hughes[1]

[1]Pearson, United Kingdom

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. The last two years have highlighted the need for flexible assessment, that can be sat at home. The days of sitting a paper-based test in a centre in an exam series 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.

It is 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.

This open paper presentation will take viewers on the journey of this standard setting process and offers a roadmap that could be used to align standards across test modes in the future.

14:45 - 15:15

Understanding Simulation-Type Items using Eye Movement, Qualitative and Log-File Data

P. Lehane[1]

[1]Dublin City University, Ireland

Many countries are now deploying online testing solutions for their terminal post-primary exams. These Technology-Based Assessments (TBAs) use items that employ a broad array of stimuli and item types. These include simulation-type test items. Although it is assumed that these can make TBAs more authentic, their impact on test-taker performance and behaviour is still unknown. While analysing the accuracy of student responses (product data) to such items is valuable, process data offers additional insights into how the responses were produced. Process data in the form of eye movements and other log-file variables were collected from 24 participants who completed five simulation-type items. Qualitative data were also gathered (n=12). Results indicated that test-takers who received full credit for their performance on these items paid more attention to relevant areas of the simulation than those who did not receive full credit. However, this finding was not consistent across all tasks, suggesting that operationalising test-takers' behaviour and performance in simulation-type items may not be as straightforward as expected. Qualitative data also uncovered the nature of test-takers' interactions with these complex items across three broad categories of behaviour: Familiarisation, Sense-making and Making Decisions.

Comparative Judgement I

13:45 - 14:15

Online moderation of non-exam assessments: is Comparative Judgement a practical alternative?

C. Vidal Rodeiro[1], L. Chambers[1]

[1]Cambridge University Press & Assessment, United Kingdom

In England, many high-stakes qualifications include non-exam assessments that are marked by teachers. Awarding bodies then apply a moderation process to bring the marking of these assessments to an agreed standard. Moderation involves checking samples of student work to ascertain whether the rank order of the work is correct and the marking criteria were applied correctly.

Comparative Judgement (CJ) is a technique where two (or more) pieces of work are compared at a time, allowing an overall rank order of work to be generated. Thus, CJ seems excellently placed to facilitate moderation.

Emerging developments in technology, allowing CJ to be implemented on digital platforms and electronic submissions of students' work, has meant that moderators can now perform the task online. This allows new ways of working and can potentially change the way in which schools/students engage with non-exam assessments.

This study explored the practical feasibility of using CJ for moderation via an experimental moderation task requiring judgements of pairs of authentic student portfolios. This included aspects such as whether moderators could view/navigate portfolios sufficiently to enable them to make and be confident about comparative judgements, on what basis moderators made their decisions, and the time taken to moderate.

Comparative Judgement Approach to marking Non Examination Assessment History: teachers' opinions

V. Rotaru[1]

[1]Qualifications Wales, United Kingdom

The use of Comparative Judgement (CJ) in assessment have been increasingly researched recently. Findings suggest that, for certain types of assessments, CJ has the potential to replace conventional marking. However, the studies are not conclusive about the potential benefits, especially for high-stakes examination environments. Moreover, research exploring the opinions of the teachers who could potentially apply CJ to high-stakes assessments is sparse if compared to other aspects, such as CJ reliability.

In Wales, many high-stakes qualifications include non-examination assessments (NEA) some of which (such as GCSE History) are marked by teachers rather than external examiners. This presentation describes a study which investigated teachers' considerations when comparatively judging NEA GCSE History essays and their views on using CJ to mark them.

In the study, 22 teachers comparatively judged 23 essays using the RM Compare™ software. Judgements were made on the same day, and during the following three weeks the judges were interviewed about their views and experience of using CJ.

The presentation will outline teachers' opinions about using CJ, and their approach to making judgements including aspects considered and approaches to making the decisions. The findings will be compared to other recent similar studies.

14:45 - 15:15

Accuracy and bias of "equating" methods based on expert comparative judgement of script quality

M. Curcin[1], M.W. Lee[1]

[1]Ofqual, United Kingdom

Standard maintaining in secondary school qualifications in England is based on statistics and expert judgement. The statistical method for setting grade cut scores involves generating outcome predictions for the current cohort based on the qualification’s historical relationship between prior attainment and outcomes. This approach cannot account for changes in cohort performance attributable to factors not in the prediction model, e.g., system-wide teaching improvement, or disruption during a pandemic. Therefore, small examiner panels additionally scrutinise small samples of examination work (scripts) around statistically recommended cut scores, making absolute judgements of performance grade-worthiness.

New methods involving comparative judgement (CJ) have been investigated which hold promise for better harnessing expert judgement and improving on the current method’s limited scope and initial dependence on statistical predictions. While CJ methods typically produce high reliability of script quality judgements, little research has examined the extent of their accuracy in replicating the difficulty relationship between test forms established via traditional equating. We report on the accuracy of outcomes from several trials and replications of different CJ methods compared to operational IRT statistical equating. We consider if any inaccuracies also reveal bias in expert judgement and discuss the potential for including CJ methods in operational standard maintaining.

Formative Assessment I

15:45 - 16:15

Changing practices in doctoral assessment: From hidden rite of passage to empowering learning experience.

D.A. Chetcuti[1], M.A. Buhagiar[1]

[1]University of Malta, Malta

Achieving a PhD is an important academic milestone that provides rite of passage into the scholarly community. This rite of a passage takes place after the successful assessment of a written thesis in a specific area of research, and in some universities the oral defence of this thesis. Due to reasons of confidentiality, assessment practices at doctoral level are usually hidden, and there is very little research about how examiners go about recommending the award of a PhD degree. The main aim of the study, which forms the basis of the proposed presentation was therefore to explore how examiners went about awarding a PhD degree, what criteria they used to make their judgement and what kind of feedback they provided for doctoral candidates. Using examiner reports from four faculties at the University of Malta covering the years 2017-2018, we analysed the feedback given by examiners. The results suggest that in their report examiners at the UoM include both summative comments based on expected outcomes of achievement outlined in the UoM PhD regulations (2008) as well as formative dialogic feedback that helped the doctoral candidates reach the required standards and enabled them to become empowered learners.

16:15 - 16:45

A further look at the impact of Covid in Wales’ personalised assessments

C. Hope[1], C. Hope[1]

[1]AlphaPlus Consultancy Ltd, United Kingdom

In this presentation we reflect on and present anonymized data gathered by the Welsh Government’s Personalised Assessments, a program of formative computer adaptive assessments introduced in 2017 and administered “on-demand” to children in Wales a maximum of any two days during the school year to learners in years 2 to 9 (Age 6 to 14). These assessments utilize the Rasch Model (Rasch, 1960/1980) to place all persons on the same ability scale across the different school years assessed.

We will present a variety of covid-gap outcomes measured by the Personalised Assessments. In each instance we will reflect on how the Personalised Assessments provide superior information to understand changes relating to the Covid-19 pandemic, and also to facilitate education more generally, compared to traditional linear pen-and-paper assessments, hopefully providing impetus for the wider use of computer adaptive assessments.

References

Rasch, G. (1960/1980). Probabilistic models for some intelligence and attainment tests.(Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.

E-Assessment II

15:45 - 16:15

Assessment and Measurement in MOOCs: Literature Review and Outcomes

A. ALLALOUF[1]

[1]NITE, Israel

MOOCs – Massive Online Open Courses – have become increasingly common in recent years. Their primary objectives are to: (1) offer learners free, or relatively low cost, access to courses; (2) use high-quality, innovative, technology-rich teaching methods; and (3) advance research on diverse topics related to learning. The review objectives were (1) to improve measurement and assessment professionals' understanding of online courses, (2) to further knowledge of the various assessment possibilities in online teaching platforms and of their implications for learning, and (3) to substantiate knowledge in the assessment of online teaching from research, academic, and professional perspectives.

Assessment of learners in MOOCs, like MOOCs themselves, can be classified according to various criteria. Several characteristics determine the type of assessment, most importantly the task type, timing, presentation of tasks, and the identity of the assessor. Online platforms facilitate implementation of methods designed specifically for MOOCs, thereby enhancing the assessment. The presentation will deal with the integration of MOOCs in the curricula of academic institutions, and with ethical considerations and methods to detect cheating. A significant outcome of the review presents the recommended type of assessment for the type of course, based on nine characteristics of online courses and learner assessment.

Extending the analysis of student performance with process data: An alternative to computer-adaptive designs in large-scale examinations?

B. Maddox[1, 2], E. de Schipper[3]

[1]Assessment MicroAnalytics, United Kingdom
[2]University of East Anglia, United Kingdom
[3]Cito, Netherlands

The digital transition has significant disruptive potential for test design, including the adoption of interactive and adaptive test formats and new uses of process data. Adaptive test designs have the advantage of being able to match the difficulty of items with respondent ability. However, in certain contexts - such as national school examinations, standardisation (rather than differentiation) may be considered as the primary basis of fairness and comparability. I.e., the idea that in each person should have the opportunity to demonstrate their ability on the same test. In this paper we therefore explore an alternative perspective, using process data from log files on respondent keystrokes and response times, as a way to extend data on student performance. The paper presents a case study of large-scale mathematics assessment in French secondary schools. With evidence from an eye tracking and video study conducted in French classrooms, and the presentation of large-scale log data, we demonstrate how ‘processes models’ can be applied to extend ‘product’ based data on test scores.

16:45 - 17:15

A Tale of Two Schools: A Case Study of Accessibility and Digital Assessment in two UK schools.

E. Barrow[1], I. Custodio[2], D. McVeigh[3]

[1]Pearson Education, United Kingdom
[2]Pearson, United Kingdom
[3]Pearson Qualification Services, United Kingdom

This presentation represents a continuation of research undertaken in 2021 that started to unpick some of the potential advantages and barriers for SEND students through the use of technology enabled assessment. The research highlighted the complex set of interconnected strands upon which the road to accessible digital assessments in predicated. Furthermore, the research provoked continued interest in deepening our understanding of how to make digital assessments more accessible, while also considering the reality of schools' current capacity to implement them. Teacher feedback indicated the presence of both institutional and practical barriers that limited the extent to which digital assessment could actually be implemented in the initial sampled UK schools. Using a case study approach focusing on two UK schools, a co-educational comprehensive school and a residential special school for physically and neurologically impaired young people, we have built on the work conducted in 2021 to develop a more detailed and nuanced understanding of accessibility of digital assessment and the SEND experience in the school and learner context. With this approach, we can illustrate the reality of the existing challenges for schools and learners of diverse and complex needs, as well as the potential opportunities afforded by accessible digital assessment.

Fairness & Social Justice I

15:45 - 16:15

The Power of Peer Assessment as an Inclusive Learning Strategy in Management Education: Fostering Social Justice at the Tertiary Level

E. Meletiadou[1]

[1]London Metropolitan University, United Kingdom

Peer assessment (PA) is one of the most popular forms of formative assessment currently used in Higher Education Institutions worldwide. As an inclusive assessment practice, PA presupposes that individual learners' needs and tastes are taken into consideration, as far as possible, to ensure that all students have an opportunity to succeed focusing on their strengths rather than their weaknesses. The overall aim is to unravel areas for improvement and supporting students throughout their learning journey. In the current study, PA was used as an inclusive assessment strategy to enhance students' writing skills and motivation towards writing. Forty-four undergraduate students attended an Academic Writing Module for one academic semester. The overall aim was to develop their writing skills taking into consideration their individual differences. Students received training in PA and were then involved in reciprocal anonymous PA. Findings indicated that learners improved their writing skills considerably and became more independent learners. Their motivation also increased as they could better understand the assessment criteria. However, they confessed that they needed more training and support in PA. The findings of the current study also indicate that PA should be initiated at a younger age to enhance student collaboration and reflection.

16:15 - 16:45

Improving High Stakes Professional Assessment with Multi-Phased Differential Item Functioning Analysis.

D. Budzynski[1]

[1]AlphaPlus Consultancy Ltd., United Kingdom

Differential item functioning (DIF) analysis aims to determine whether any demographic groups perform worse (or better) than expected on each item in an assessment, but knowing where DIF occurs does not provide an answer to what the root causes of DIF are, or what to focus on when writing new items to avoid it as much as possible.

AlphaPlus worked with the Bar Standards Board and developed a novel multi-phased approach for conducting DIF analysis. It consists of a classic DIF analysis, focus groups with experts helping identify reasons for DIF, and a series of regression analyses aimed at establishing whether there are particular item features (including those flagged in focus groups) which tend to be associated with DIF.

This method allowed us to not only identify which items exhibited DIF, but it also flagged many factors to consider when reviewing or writing items to mitigate DIF as much as possible.

The presentation will explain our novel DIF analysis approach, its theoretical underpinnings, and key things to consider when carrying out this type of analysis.

16:45 - 17:15

What should a coherent and inclusive qualification offer, for 14-16-year-olds, look like in Wales? How can it prepare them for life after 16?

J. Moriarty[1]

[1]Qualifications Wales, United Kingdom

Learners of tomorrow will need to have an adaptable skill set when entering the world of work with jobs that do not exist yet. Stakeholders have emphasised the need for young people in Wales to feel ready and prepared for post 16 life, including the development of skills needed for work and employment. It can be questioned; are qualifications needed to ensure young people have developed these required competencies?

The challenge is then to ensure that there are qualifications for learners at all levels to access, that they see relevant to their lives. Are the development and assessment of skills in readiness for the world of work the glue to make this happen?

Ensuring there are accessible qualifications that will allow for progression to a wide range of pathways is a fundamental expectation. In Wales, significant curriculum reform is taking place and the 14-16 qualification offer is a crucial piece of the jigsaw.

Qualifications Wales' review of the wider offer of qualifications at 14-16 explores what is needed in a future offer to be inclusive and coherent and identifies skills learners of the future will need to progress. It considers the role of assessments, digital technology, and other emerging considerations.

Assessment Cultures I

15:45 - 16:15

Students’ engagement in policy experimentation: views and experiences of the (mis)use of digital formative assessment in schools across Europe

J. Elwood[1], K. Livingston[2]

[1]Queen's University Belfast, United Kingdom
[2]University of Glasgow, United Kingdom

This paper presents ongoing analyses of data from the Assess@Learning study. Two seemingly disparate areas of educational interest that are at the foundation of the study are brought together in our presentation: assessment reform and student voice. The policy experimentation study foregrounds the experiences of formative assessment and the use digital tools in school to enhance students’ learning. Our focus is predominantly on data gathered from 122 students in 2021 to provide an insight into their lived reality of experiencing formative assessment and in particular, digital formative assessment. A qualitative research approach was utilised to capture rich in-depth dialogue during Student Dialogue Labs (student only, facilitated by adults), and Country Dialogue Labs (students and adults in dialogue) held in 5 European countries participating in the study. The emergent themes from our analysis of dialogue will be shared to highlight those aspects of digital formative assessment that students give attention to and which they see are pertinent to improving current experiences. The Assess@Learning study provides a unique opportunity to look more systematically at the digital transformation of formative assessment as well as to have students as key informants on the efficacy of such practices as they experience them in educational settings.

What contributes to success in GCSEs in England? A predictive validity analysis and a model of holistic understanding

I. Suto[1], T. Benton[2], G. Copestake[1]

[1]Cambridge Centre for Evaluation and Monitoring, United Kingdom
[2]Cambridge University Press & Assessment, United Kingdom

The pandemic has caused students to miss opportunities for both academic and social learning. Following lengthy closures, England's schools have reopened, but in these uncertain times, the need for a more holistic teacher approach to understanding and supporting students has become more salient. We explore what this means for the assessment community.

First, we analysed the predictive validity of some assessments of cognitive and cross-curricular skills. The assessments were taken almost six years prior to the students sitting GCSE examinations (N = 9,967), and correlation coefficients were found to be high: 0.77 for GCSE Mathematics, 0.72 for English, and 0.71 for Geography. Predictive validity levels were at least as high as for national curriculum assessments.

Secondly, since our analysis revealed high but imperfect predictive validity levels for cognitive skills, cross-curricular skills, and precursor curriculum coverage,
we explored what had not been assessed. A literature review revealed two additional areas of teacher insight: personal attributes such as wellbeing, and students' environments. The five interacting areas of insight form a holistic model. We use it to discuss how teachers could combine numerical data from baseline and formative assessments with observational, student discussion, and qualitative judgement data, to build actionable student profiles.

16:45 - 17:15

UDL and Inclusive Assessment for University Students with Intellectual Disabilities

D. Camedda[1], J. Banks[1], M. Shevlin[1]

[1]Trinity College Dublin, Ireland

Increasingly diverse student populations have been a key factor in the development of inclusive practices and policies within third level institutions over the last two decades. As part of this, the effectiveness of how institutions teach and assess this diverse student population has been examined. Universal Design for Learning (UDL), is an instructional approach designed to make curriculum and learning more accessible for every student (Boothe et al., 2018) through three main areas: engagement, representation, action and [removed]Meyer et al., 2014).

This paper presents how UDL principles have been applied to the assessment framework within a two-year Certificate in Arts, Science and Inclusive Applied Practice, a post-secondary programme for students with intellectual disabilities in Trinity College Dublin. It details the benefits and challenges of the application of multiple means of action and expression in assessing the learning of these students throughout the course. The paper also describes how a survey and focus group will be conducted with the students on the perceived effectiveness of the UDL principles applied to assessment. The results of this research are to be presented at the AEA conference.

National Tests & Examinations I

15:45 - 16:15

Taking a fresh look at assessment: why we must modernise our system of assessment to realise the full potential of digital technologies

D. Seabrook[1]

[1]Qualifications Wales, United Kingdom

This paper proposes that the full potential of digital technologies to transform assessment will only be possible by taking a fresh and foundational look at the assessment system overall. Through a series of semi-structured interviews and focused discussion groups, Qualifications Wales has found that digital technologies are having a transformative impact on teaching and learning, which has been accelerated by the necessity of digital transition during the Covid-19 pandemic, presenting a challenge to the assessment community to ensure it is suitably aligned.

The tools available to educators are enhancing learners’ agency in their development, encouraging participation and extending learning beyond the classroom. The repertoire of assessment tools is continuing to expand and digital assessments provide more ways than ever to capture what learners can do. Yet without rethinking the structure of our assessment system – from our conception of reliability to the way assessment is scheduled – assessment, digital or not, will be bound by the past and unfit to measure what it must in the future.

Qualifications reforms: opportunities and challenges - a focus on A level Mathematics in England

G. Grima[1], B. Redmond[2], J. Golding[3]

[1]Pearson UK, United Kingdom
[2]Pearson, United Kingdom
[3]University College London Institute of Education, United Kingdom

In the context of a rapidly changing educational landscape and fundamental shifts in the skills required for employment, we present an overview of the challenges and opportunities of qualification reform. Using data from a four-year study (2017-2021) which explored reformed Mathematics A levels in England, we discuss the opportunities for more aspirational and holistic learning created, aligning with the needs of higher education and future employers. However, we also show challenges for teachers, already under pressure, as they adapt to changing expectations while engaging with more mathematically demanding content and teaching for increasingly high-stakes assessments. We also draw on 'The Future of Qualifications and Assessment in England' report which shows that teachers, while supportive of reform in principle, are cautious about impacts of successive reforms in short timeframes . When considering post-pandemic reflections, we saw appetite for more diverse forms of assessment, but there was also recognition from a range of stakeholders that high-stakes assessments have a central role to play in the educational journey of learners. In line with this, feedback from multiple stakeholders put emphasis on evolutionary change over time as a sustainable model for improving educational systems, building on current strengths, utilising institutional memory, and maintaining stability.

16:45 - 17:15

The digital transformation of teaching and learning for high-stakes assessment: teacher and student responses in England.

B. Redmond[1], J. Golding[2], G. Grima[3]

[1]Pearson, United Kingdom
[2]University College London Institute of Education, United Kingdom
[3]Pearson UK, United Kingdom

We discuss the use of digital technologies in reformed Mathematics A levels in England. We draw on evidence from a four-year study (2017-2021) exploring implementation of these qualifications. Pre-pandemic findings typically showed modest use of digital resources in teaching, although where utilised, teachers and students were positive about their potential for enhanced mathematics learning. Teachers' own limited experiences with technology and digital pedagogies, poor access to devices, and poor assessment reward for such activity, created disincentives for use, resulting in the marginalisation of key areas of the intended curriculum. Pandemic-driven school closures saw participants rapidly adjusting to online learning, with accompanying digital tools primarily supporting communication, organisation and administrative purposes rather than subject-specific functions such as graphing or data handling. Early, significant challenges with teacher preparedness and student access to technology gradually improved and by Summer 2021, comparatively few students were experiencing serious access constraints, and teachers were developing confidence and fluency in using digital tools. However, in most cases the use of digital tools for specifically mathematical purposes stagnated or deteriorated. Since subject-specific use of digital tools can enhance learning in multiple powerful ways, it is important to find ways to support teacher and student adaptation to these also.

Assessment Against the Backdrop of Covid II

15:45 - 16:15

Reflections on teacher assessment after the 2021 Teacher Assessed Grades process in England

S. Vitello[1], T. Leech[1]

[1]Cambridge University Press & Assessment, United Kingdom

Greater use of high stakes teacher assessment is being increasingly considered as an alternative to exams. This has been given further impetus by the experience of the Covid-19 pandemic, which in England saw 2020 and 2021 exams replaced with teacher assessment. In 2021, schools were allowed substantial latitude to determine candidate grades from a broad variety of available evidence. We report an analysis of candidate evidence submitted to the OCR exam board by schools to support 2021 Teacher Assessed Grades in English language and maths GCSE qualifications. We consider the types and volume of evidence selected by schools, the conditions under which it was produced, and how different evidence was prioritised. We found that teachers tended to use evidence from exam style assessments such as past papers, perhaps due to its more standardised nature. Many preferred assessments taken under exam conditions where possible. English language saw slightly more non-standard assessments including additional writing tasks, but materials similar in form to those of a normal exam-based session were common. We consider implications of these findings for designing high stakes teacher assessment in future, non-emergency situations to be practicable from the teacher perspective and to ensure rigour and comparability of standards.

How Swedish schools, universities and large-scale assessments were affected by COVID-19 and what we learnt

M. Wiberg[1], P. Lyrén[1], A. Lind Pantzare[2]

[1]Umeå University, Sweden
[2]Umea university, Sweden

The overall aim is to describe, analyze and discuss how COVID-19 affected Swedish schools and the national tests in schools, university teaching and examination, and the college admissions test, Swedish Scholastic Aptitude Test (SweSAT). Another aim is to discuss challenges in college admission processes, universities, and schools due to COVID-19. During the COVID-19 pandemic, Swedish schools, except at the upper secondary level, remained open in Sweden during the whole pandemic. Higher education institutions basically closed their facilities in March 2020 and instruction went online, while upper secondary schools was semi-open. The college admissions test SweSAT and the national tests for grades 3, 6 and 9 were however cancelled, which had impact on students' application to college as well as on the students' grades. In this presentation we discuss, describe, and analyze the events by using documentation from news, school and university authorities, governmental reports and through a student survey. We also reflect on the possibilities and challenges and what we can learn from the pandemic.

16:45 - 17:15

The post-pandemic comparability narrative. What changes might we expect?

G. Elliott[1]

[1]Cambridge University Press and Assessment, United Kingdom

Comparability has been an issue for providers and users of assessments in England for more than a century. Integral to the fairness and justice agenda, the end users of our assessments require us to have a coherent, stable and well-understood mechanism for the parity of different qualifications pathways. How then, have the enforced events of the past two years affected these mechanisms and what should we be doing about it?

In this piece, I shall explore the comparability challenges likely to be faced in England over the forthcoming ten years and compare these with some of the issues from the past. Do we face greater comparability issues than ever before, new comparability challenges, or does the future merely hold a further iteration of the shifting public focus from one type of comparability to another?

Looking specifically at the needs of the immediate generations of students to come and taking heed of some elements of hindsight from the past, what are the actions that we, as assessment professionals, need to take in order to reassure our customers and stakeholders?

Assessment of Practical Skills I

15:45 - 16:15

In occupational assessments do we always have to assess all learning outcomes?

A. Boyle[1], H. Limmer[2]

[1]AlphaPlus Consultancy, United Kingdom
[2]AlphaPlus Consultancy Ltd., United Kingdom

Occupational and work-related assessment is an important part of the education landscape, and yet is relatively undertheorised. In this paper, we look at how component results (question, unit or module scores) are aggregated to provide overall qualification or examination outcomes.

In particular, we focus on conjunctive aggregation, (or ‘multiple cut off’ models), which is when a criterion must be met regarding each source of information separately; that is, separate standards exist for each source of information. This approach – often contrasted with compensation – is common in occupational assessments. We report a synthetic literature review, which draws out ideas behind conjunctive aggregation, and which questions assumptions underlying this approach.

Then, we go on to look at three UK assessment systems, and how they aggregate scores, and the extent to which they embody a ‘classical’ conception of conjunction, compensation, etc..

We aim to dig deep into the roots of a seemingly straightforward assessment issue – ‘how do you combine scores?’ – and provide some useful insights into assessment theory and practice, the nature of professional knowledge and expertise, and even how occupational assessment interacts with the demands of the labour market.

The assessment of behaviours in apprenticeship end point assessments

D. Tonin[1], T. Aston[1], L. Clarke[1], S. Cadwallader[1]

[1]Ofqual, United Kingdom

In England, End-Point Assessment (EPA) is the final stage of an apprenticeship, and it is designed to test whether apprentices are fully capable of doing their job. During their EPA, apprentices must demonstrate the knowledge, skills and behaviours (KSBs) that employers identified as important for that role.
Besides technical knowledge and skills, behaviours (also called competencies or soft skills) are essentials for future employees. Behaviours, including communication abilities, problem-solving, critical thinking, presentation skills, leadership, and teamwork amongst others, are more challenging to assess compared to knowledge and skills because they may be demonstrated only under specific circumstances or sporadically.
Ofqual, the regulator of qualifications in England, has recently started providing external quality assurance (EQA) for apprenticeships. To deepen our understanding of EPAs and therefore inform our approach to regulation, we explored how behaviours are assessed in EPAs using qualitative and quantitative methodologies. The presentation will discuss how behaviours are currently described in EPAs (given that behaviours are not universally described across job sectors), which assessment methods are currently used to assess behaviours and how the approaches to assessment vary by the target behaviour and by job sector.

Psychometrics and Test Development II

15:45 - 16:15

A general framework for measurement applied to temperature and reading

J. McGrane[1], A. Maul[2], D. Briggs[3]

[1]University of Oxford, United Kingdom
[2]University of California - Santa Barbara, United States
[3]University of Colorado - Boulder, United States

Across the scientific and philosophical literature, it is difficult to identify a single definition of measurement applicable across all disciplinary contexts, and moreover, it is not clear that this is even a valuable goal. However, given the widespread trust afforded to measurement processes and their results, we think it would be valuable to identify what elements of a measurement process serve to justify this trust, independently of the domain of application. With this goal in mind, we propose that a measurement process (a) is expected to provide information about an attribute of an object or event in the world (the target attribute), (b) uses instrumentation designed to be sensitive to differences in the target attribute, and (c) produces information in the form of standardised values of that attribute, which is in turn (d) qualified by information about the uncertainty of the values. We take these four elements -- substantive theory, instrumentation, standardisation, and mathematical analysis (TISM) -- as domain-independent characteristics of measurement processes. We argue that viewing measurement processes in light of such a meta-theoretical framework provides a basis for seeing what is common to measurement processes in domains of application as diverse as the physical sciences and education.

16:15 - 16:45

Detecting and Responding to Malpractice in on-line and automated scoring systems

R. Clesham[1], S. Hughes[1]

[1]Pearson UK (corporate membership), United Kingdom

While most test-takers approach tests in good faith, assessment systems must find ways to detect and prevent malpractice where it happens. Regardless of assessment mode, malpractice is always a potential threat to validity.

Since their inception, the integrity of high stakes test outcomes has required systematic organisation and processes. In the UK, the Joint Council for Qualifications (JCQ) provides a central unified role to ensure that public examinations are conducted with due regard to examination integrity. In addition, Awarding Bodies carry out a number of internal procedures and processes in order to respond to allegations of malpractice. However, in general, dealing with malpractice has not changed for centuries.

The size, scale and operation of globalised high stakes assessments expands annually. Tests and examinations are increasingly computer based and marking is either carried out by automated systems or through on-line human marking where whole scripts are not seen. This presentation will describe how automated monitoring and scoring systems aggregate and synthesise enormous volumes of data in order to detect patterns of malpractice/gaming behaviour that would largely remain undetected in more traditional human-scored assessments.

Performance-based measurement of communication and cooperation skills: standardized test results vs. role play assessment

A. Granaturova[1], I. Uglanova[2], M. Lebedeva[3]

[1]Higher School of Economics, Russia
[2]National Research University Higher School of Economics, Russia
[3]Pushkin State Russian Language Institute, Russia

Assessing soft skills such as communication and cooperation is one of the most ambitious goals in modern education. The aim of the research is to present the validity evidence of computerized performance-based assessment (CPBA) of communication and cooperation skills in secondary school. The CPBA includes six scenario-based tasks that mimic real-life situations. To provide evidence of ecological validity the qualitative research in the format of the role play was conducted. The students (26 sixth graders) were asked to play roles that were described in the scenario. The key behavior indicators were assessed similarly to the scoring rules of the CPBA. After that, the students were asked to solve the computerized task. The results demonstrated that similarity between students' behavior within the CPBA and real-life communication was found, however, significant features of real communication were not captured by the standardized CPBA. To provide factor validity and fairness of assessment, we conducted a quantitative data analysis of 1680 students' profiles (6, 7, and 8 grades). Communication and cooperation skills demonstrated unidimensionality based on the results of CFA. Measurement invariance across grades was partially achieved. The study reaffirms the advantages of computerized assessment of complex skills in classroom settings.

Friday,
11 Nov, 2022

Assessment Against the Backdrop of Covid III

9:00 - 9:30

The Impact of Pandemic-Necessitated Adaptations of a Selection Process (MMI and Questionnaires) Used in Medical School Admissions

D. Ziegler[1], A. Moshinsky[1], G. Soffer[1], E. Ben-Barak[1], D. Gilon[1], D. Sagi[2], K. Macmillan[2], A. Ziv[2]

[1]NITE, Israel
[2]MSR, Israel

Most Medical schools in Israel adopted assessment centers that include behavioral stations and questionnaires as part of their admissions process.
In 2020, after the outbreak of the COVID-19 pandemic, medical schools requested that adjustments be made in order to administer a comparable version of the assessment center while complying with Covid-19 restrictions. Accordingly, the following adaptations were made:

- Two of the eight behavioral stations were replaced by a computerized personality questionnaire.
- Most interview stations were conducted remotely via Zoom, but few were still conducted in person.
- Simulation stations in which actors and candidates are engaged in sensitive interpersonal interactions would be compromised as an evaluation tool, if they were conducted remotely, were still administered in person.

Feedback questionnaires from both evaluators and candidates indicated that the adaptations were successfully implemented.
Our presentation is of data pertaining to the scores of candidates from the two years during which the selection system was administered in compliance with Covid-19 restrictions. We also provide data from candidate scores in 2019, the year preceding the pandemic. We compare interview stations administered remotely with those administered in person, and simulation stations involving wearing masks with those in which this was not the case.

9:30 - 10:00

Transition towards item response theory based standard setting in medical assessments

M. Turner[1], B. Smith[1]

[1]AlphaPlus, United Kingdom

Getting standard setting right is critically important, especially in medical assessments. But many professional assessments standard set in a very different manner to the arguably ‘lower stakes’ school exams many assessment professionals are chiefly involved with.

An example of such a medical qualification is the General Pharmaceutical Council (GPhC)’s, registration assessment which historically used a modified Angoff method was used to set the pass score. However, GPhC encountered ‘creep’ in the rating of items over time, which would have led to the pass mark being inflated without intervention. This issue can occur in any approach that repeatedly sets a new standard each year, incentivising the use of a statistical standards maintenance approach – but such approaches are relatively rare in medical assessment.

AlphaPlus has worked with GPhC since 2016 and helped move the registration assessment to a standards maintenance approach in 2020, avoiding the issues caused by needing to re-set the standard judgmentally each year. This presentation will detail how this transition was successfully implemented, focusing in particular on the challenges faced and how these were overcome, with a view to providing a good case study for shifting to a standards maintenance model other professional assessments could make use of.

Student perceptions of the adaptations to exam arrangements in England in 2022

F. Leahy[1], M. Curcin[1], A. Brylka[1]

[1]Ofqual, United Kingdom

The uncertainty and disruption caused by the pandemic has potentially increased the pressure students feel about revising for, and sitting, exams. A package of adaptations to exam arrangements was implemented in England in 2022 in order to mitigate some of the impacts of this disruption. One of the main adaptations was that advance information about topics that would come up in many national exams was published approximately 3 months before the assessments took place. The aim was to help students focus their revision, and to reduce some of the anxiety students may feel about their exams. This is the first time such information has been released in advance of national exams. Other adaptations included a reduction in content in some subjects, and the permitted use of formulae sheets in some maths and science exams. This study aimed to understand student perceptions of these adaptations, primarily advance information as this is a novel assessment arrangement. This will be important should the use of such adaptations be considered again in future. Results of an online survey of students taking exams in 2022 will be presented, alongside the findings from follow-up focus groups.

Education & Policy Assessment II

9:00 - 9:30

Patterns of educational inequalities in mathematics and science: An analysis using three cycles of TIMSS in Ireland

A. Duggan[1], A. Karakolidis[1], A. Clerkin[1], L. Gilleece[1], R. Perkins[1]

[1]Educational Research Centre, Ireland

Socioeconomic characteristics are persistently and systematically related to academic outcomes, despite efforts to reduce educational inequality. This study uses data for Ireland from the Trends in International Mathematics and Science Study (TIMSS) to investigate patterns of socioeconomic inequalities in 4th grade students' performance in mathematics and science from 2011 to 2019. Two measures of inequality are examined: (i) inequality of achievement, i.e., the degree of variability in student performance and (ii) inequality of opportunity, i.e., the extent to which student performance is shaped by demographic and socioeconomic characteristics.

Descriptive and multilevel regression analyses are conducted to explore variability in student performance and to investigate the variance in achievement explained by school- and student-level socioeconomic factors, across cycles and subjects.

Findings indicate that significant improvements observed in average achievement are not necessarily accompanied by reduced inequality. Between 2011 and 2015, inequality of achievement decreased; however, between 2015 and 2019 a small increase in inequality of achievement was observed. Inequality of opportunity has increased steadily since 2011. Findings were consistent for both subjects. This study provides valuable findings in the context of measuring and tracking educational inequalities.

9:30 - 10:00

A Policy Document Analysis of Post-Soviet Assessment Policy in Kazakhstan

R. Kakabayeva[1]

[1]King's College London, United Kingdom

In 2015, a gradual transition to a new criteria-based assessment policy consisting of formative and summative assessment models began in mainstream schools in Kazakhstan. From the first day of its introduction, the new assessment policy has been a controversial and much disputed subject within society. This study provides an overview and analysis of the assessment policy documents underpinning criteria-based assessment policy. The study outlines educational assessment policy in Kazakhstan through the analysis of a wide range of documents that have contributed to the history, development, implementation, and evaluation of the new assessment policy with the aim of exploring the links between these and the policy imperatives framing new visions of assessment.

Document analysis revealed a very detailed image of social, economic, political, and ideological shifts as well as the origins and growth of new assessment visions, including their aims, methods, and desired outcomes, as well as their applicability to school life. Similar thoughts were extracted from a document analysis conducted by newly created quasi educational officials, which emphasised the creation of favourable conditions in schools for the implementation of new evaluation methods. Following that, document analysis allowed for a deeper dive into the many viewpoints on the new assessment system.

10:00 - 10:30

A Tool for Measuring 21st Century Skills in Qualifications

G. Mann[1], E. MacIntosh[1]

[1]SQA, United Kingdom

The pandemic has made it clear that young people can adapt to societal disruption, are resilient and have the capacity to learn creatively in unpredictable situations. They have shown that, when given the opportunity, they can embody 21st century skills and are open to new and creative learning, teaching and assessment approaches. As assessment professionals, we must continue to build on these gains by designing qualifications and assessments that support 21st century knowledge and skills development so that young people can flourish in an unpredictable world.

The intention of Scotland’s Curriculum for Excellence is to build capacity for 21st Century thinking by developing skills for learning, life and work. However, understanding the extent to which qualifications and assessments meet the intentions of Curriculum for Excellence has been an issue. It is vitally important that we not only aim to develop these skills through our qualification design, but also have a mechanism for measuring whether they are being effectively developed in practice. As researchers at the Scottish Qualifications Authority (SQA), we have developed a methodology for measuring whether our qualifications meet the intentions of Curriculum for Excellence that can be adapted to measure 21st skills in qualifications more generally.

National Tests & Examinations II

9:00 - 9:30

Links between school grades, standardized tests and gender: fairness of assessment from resilient students' perspective

E. Melnikė[1], D. Sevalneva[1], R. Erentaitė[1]

[1]Kaunas University of Technology, Lithuania

Current challenges during COVID-19 pandemic activated discussions and even actions for more rapid shift from external standardized assessment to more emphasis and trust to school level/teacher assessment. A lot debate is going on about which is better for todays and future students learning. Assessment is an integral part of teaching and learning process, so no matter what the method or form of assessment, its primary purpose is to help student in his learning process, providing valuable and reliable information. The results of international surveys show that students from socioeconomically disadvantaged homes on average demonstrate lower academic achievement. However, research also shows, that there is a group of disadvantaged students who achieve the same high results as their more advantaged peers (Agasisti et al., 2018). This capacity is called academic resilience.

The purpose of this study is using population-based data to investigate the links between school grades, standardized tests and gender with special focus to disadvantaged students.

Study results shows that among resilient students, a difference between boys and girls results in math grades (school level assessment) were much bigger than in math test scores (external standardized assessment).

9:30 - 10:00

How do school systems respond to examinees who experience illness or bereavement at the time of exit examinations in secondary education?

D. Murchan[1], E. Likhovtseva[1]

[1]Trinity College Dublin, Ireland

This study investigates how high-stakes examination systems respond when students are unable to present for scheduled assessments. Unexpected illness and injury or the effects of the bereavement of a family member close to or during exams can introduce a potent dual challenge to students. This involves dealing with the personal trauma and with the fallout of missing a test that carries significant consequences for students' next steps in education or employment.

The aim of the study was to investigate approaches used in high-stakes examinations systems to meet the needs of students who experience such traumatic events during examination periods. A mixed-methods approach explored the variables of interest through three lenses: (i) existing academic literature, (ii) policy and protocols published in agency websites and (iii) case studies in a sample of education systems, involving interviews with key officials.

Findings draw on data from 19 education systems, revealing how education systems conceptualise the problem and respond, often in diverse ways. The findings provide insight into an assessment issue that is modest in terms of the overall population of students but hugely significant to the small minority of students who find themselves unexpectedly unable to travel to the exam hall.

10:00 - 10:30

Longitudinal relationships between school achievement, self-beliefs and mastery goals over Grades 6-12

H. Eklöf[1], A. Hofverberg[1], E. Knekta[2]

[1]Umeå University, Sweden
[2]Umeå Universitet, Sweden

The present study aimed to investigate whether students' self-beliefs and mastery goals have an impact on later achievement over and above that of previous achievement and cognitive ability, and also whether there are signs of reciprocal relationships. To achieve this aim, a longitudinal data set with 2,256 students assessed at three time points over six years was used. National test scores in math and the Swedish language and more general tests of verbal and quantitative ability were included in the analyses as were measures of students' self-beliefs and mastery goals. An autoregressive cross-lagged panel path model with control variables was used to investigate relationships between the different variables over time. Findings showed that earlier motivational beliefs predicted later motivational beliefs, and earlier performance predicted later performance, but also that motivational beliefs may affect later performance and vice versa. This suggests that the skills and knowledge the students bring to school need not be decisive of later performance, but that motivational variables can be significant predictors over and above that of previous performance. Findings may have implications for how student motivation and attitudes should be acknowledged in everyday learning situations and in a changing assessment system with changed assessment practices.

Higher Education & Assessment

9:00 - 9:30

Modeling examinee answer changing on multiple choice tests

E. Papanastasiou[1], A. Stylianou-Georgiou[1]

[1]University of Nicosia, Cyprus

Due to the latest technological and psychometric advances related to the field of testing, coupled with the disruptions that have occurred in education due to the COVID-19 pandemic, many organizations have moved towards the online testing of students. As a consequence of this change, process data from log files have become easily accessible to researchers. Through the analysis of such data, researchers and measurement specialists can gain more insight into both, examinee abilities, as well as item and test quality. What log files lack however, are the examinees' opinions on why they performed certain test-taking practices. Due to this need, and the importance of also understanding how examinees approach test-taking, this study attempts to model the interrelationships among test-taking strategy instruction (with a focus on metacognition), answer changing bias and performance on multiple-choice tests among college students. The data for the study were obtained from a sample of 1512 students from Greece, which were analysed with the use of structural equation modelling. This study manages to extend the findings of previous research by proposing a model that demonstrates the interplay between answer changing, answer changing bias, testing strategies utilized by students, and their relationship to overall achievement.

9:30 - 10:00

A New Model for Critical Thinking – Addressing the Growing Global Skills Gap

G. Hudson[1]

[1]GA Partnership, United Kingdom

Creativity and Critical Thinking Skills in Schools is high on the agenda of the OECD and its conference of the same name held in London in 2019 focused solely on this important topic. It is now widely recognised that students of today should be equipped better to deal with the significant changes in the future working environment. Jobs that have relied upon knowledge and experience and skills that decay over time will be replaced by those that depend more upon the harnessing of augmented intelligence. Teaching critical thinking strikes to the heart of this.

This presentation will describe a research-based critical thinking model developed over the past ten years by MACAT – and how it is being applied now in university settings globally. The PACIER model comprises six component skills (Problem Solving, Analysis, Creative Thinking, Interpretation, Evaluation and Reasoning) plus twenty-four sub-skills. Details of the ecosystem that makes the MACAT approach successful will be given explaining the framework (PACIER), the process (test, teach and track) and the supporting materials (assessments, a development library and taught masterclasses). Its compelling relevance to the future global economy will be referenced.

10:00 - 10:30

University Students' Perspectives Towards Online Assessment in the COVID-19 Era

A. Keane[1], K. McFerran[1], B. Acton[1], S. Taylor[1], D. McLaughlin[1]

[1]Queen's University Belfast, United Kingdom

During the COVID-19 pandemic, educational disruptions necessitated Higher Education Institutions (HEIs) to reimagine their pedagogical approaches with lecturers required to translate campus-centric teaching and assessment material into online formats. With the gradual return to on-campus activities, the potential of digital affordances in traditional student-facing institutions is currently being debated.

The primary purpose of this qualitative study was to explore the online assessment experiences of a group of undergraduate Health and Life Science students. Data was collected by interviews and focus groups. Each session was audio-recorded, transcribed and analysed to identify descriptive categories of students' conceptions of assessment.

From the students' perspective, individual and group oral presentations did not accurately capture effective communication or teamwork skills when conducted online. The data further highlighted students' identification of timed closed book examinations as tests of memory recall. Instead, students expressed a preference for open book examinations which allowed them to critically apply their knowledge.

The general impact and efficacy of digital technology in Higher Education since the start of the pandemic requires considered evaluation for its putative support for teaching and assessment. An improved understanding of the online assessment experience will help inform improved pedagogies and assessment in blended and online learning environments.

Assessment Cultures II

9:00 - 9:30

An analysis of contrasting approaches to the assessment of Technical and Vocational Education and Training (TVET): cultural insights from practice in England and Germany.

E. Andressen[1, 2], S. Shaw[2]

[1]Andressen Byram Ltd, United Kingdom
[2]University of Cambridge, United Kingdom

Assessment practice is rooted in the culture in which it is located. This research addresses the question of whether an understanding of transnational assessment cultures can inform local assessment practice. The research aims to demonstrate an understanding of why there are differences between the two assessment cultures and what they mean for the structure, practice and use of TVET in local contexts. Apprenticeships are available to young people aged 16-19 in England, but there is still a positive bias towards academic study leading to university. Multiple roles are assigned to vocational study in England, such as qualifications reform as a proxy for system reform, and a ‘non-academic’ alternative, which nevertheless serves to qualify for, and route learners into, Higher Education. An apprenticeship at 16 in Germany is seen as a positive choice for many who are not continuing the academic route to Abitur. Additionally, two programs in Germany prepare young people for vocational training if they have been unable to find an apprenticeship placement on completion of their Schulabschuluss. This presentation will look at what can be learned from the two assessment cultures. We consider elements such as attitudes towards TVET in society, social partnerships, modularisation of learning, and employability.

Assessment for Learning practices and teachers' assessment culture: what synergies?

R. Pasquini[1], F. Morales Villabona[1]

[1]University of teacher education Vaud, Switzerland

Since 2019, the Ministery of Education from the State of Vaud, in Switzerland, promotes Assessment for Learning (AfL) practices. Our exploratory study aims to examine how secondary teachers have implemented AfL practices in their classroom, and to better understand the role their assessment culture mindset (Birenbaum, 2016) played in this process. Data were derived from six secondary teachers (Science and French) involved in a one year collaborative research. Various qualitative data were gathered: interviews, writing tales, working session recordings and teachers' artefacts. In order to describe teachers' assessment culture mindset, we used conceptualizing categories (Paillé & Mucchielli, 2012) related to five AfL principles: using assessment criteria, establishing synergies between formative and summative assessment, giving feedback that support learning, designing complex assessment tasks and using learning objectives. Our findings document the complexity of the AfL practices' development and also some issues teachers faced. We put in light that teachers' assessment culture and the mindset it reflects depends of a lot of factors such as the teaching subject or the theoretical model proposed during the working sessions. Obviously, further research should deepen the conceptualization of assessment culture, in order to better understand its dynamic dimension and its role in AfL implementation.

Assessment of Practical Skills II

9:00 - 9:30

Assessment of scientific literacy to predict student's ability to study science-oriented subjects: Case of Nazarbayev Intellectual schools

A. Zhussupov[1], A. Jandarova[2], L. Tursynova[1], A. Zhapparova[1]

[1]NIS CPM, Kazakhstan
[2]AEA NIS, Kazakhstan

Scientific literacy in secondary education is associated with “Natural science” related subjects such as Physics, Chemistry, Biology, Geography. In Kazakhstan, within the large-scale update of national secondary curriculum, the conceptual framework of “Natural Science” subject was reconsidered to develop students' basic knowledge and skills.

One of the leading selective school networks in Kazakhstan, Nazarbayev Intellectual schools (NIS), in 2021 in collaboration with Institute of Pedagogical measurements, Netherlands (Cito) incorporated new subject test “Natural Science” to its student's selection system. The new test section is aligned with updated “Natural Science” curriculum for grades 5 and 6.

In this study we examine whether the assessment of applicants' scientific literacy through “Natural Science” subtest helps to identify their abilities for successful learning of Physics, Chemistry, Biology, Geography at NIS Grade 7.

For this purpose, the results of the approbation test, which was conducted during selection test in 2021 and the academic performance in “Natural Science” related subjects in Grade 7 during the 2021-2022 academic year of 2 976 students will be analysed using multivariate models. Also, “Mathematics” and “Natural Science” subtests results will be examined in terms of their interrelation by employing confirmatory factor analysis.

9:30 - 10:00

Assessing creativity among primary school children through the snapshot method – an innovative approach in times of uncertainty

S. Johnston[1], T. Denton-Calabrese[1], J. Scott-Barrett[1], J.A. McGrane[1], T.N. Hopfenbeck[1]

[1]The University of Oxford, United Kingdom

Creativity is an essential cognitive and practical skill for life-long learning and its assessment has been the focus of several international debates. However, the typical approaches of assessing creativity are often labour intensive, and therefore, time consuming. This study evaluates the snapshot method, including its psychometric properties, of assessing primary-aged children's (n = 179; M age= 9.10, SD = 0.90) creative responses generated within an alternate uses task. The creativity task yielded 534 responses, and together, three raters provided a total of 1,602 ratings. Using a 1 to 3 rating scale (1 = Not at all creative, 2 = quite creative, 3 = highly creative), responses were rated according to a rubric developed for this study. Through confirmatory factor analysis (CFA), the results showed that generally good construct validity and reliability can be achieved using the snapshot scoring method. Together, the findings suggest that snapshot scoring appears to be an efficient and promising approach to assessing creativity among primary-aged children. Recommendations are offered for researchers aiming to employ this method among school-aged children, including broader discussions around rating scales and rubrics within the snapshot scoring method.

Examining the generalizability of process indicators for planning and non-targeted exploration and their relationships with problem-solving competency

M. ZHANG[1], B. Andersson[1], S. Greiff[2]

[1]University of Oslo, Norway
[2]University of Luxembourg, Luxembourg

Problem-solving competency is essential in modern society. We focused on two important cognitive processes in problem solving, planning and non-targeted exploration. With log files from the problem-solving domain of 2012 Programme for the International Assessment of Adult Competencies, we derived process measures for planning and non-targeted exploration. Next, we examined whether the process indicators could be generalized across multiple tasks using confirmatory factor analyses. We finally investigated the relationships between planning, non-targeted exploration, and problem-solving competency on the latent and observed variable level via multidimensional latent variable analysis. The results indicated that (a) the planning indicator was well generalized across tasks, whereas the non-targeted exploration indicator was less generalizable under the full model; (b) non-targeted exploration was strongly related to problem-solving competency, whereas the correlation between planning and competency was slightly negative on the latent variable level; and (c) such relationships on the observed indicator level varied substantially across tasks. Different from most studies analyzing process data from single tasks, we expanded the focus to multiple tasks and connected the observed indicators with relevant latent traits. The task-dependent relationships deepened our understanding that planning and non-targeted exploration functioned differently in complex problems, requiring careful examinations when generalizing conclusions among tasks.

Formative Assessment II

9:00 - 9:30

Linking Standardized and Formative Assessments: A Predictive Perspective

S. Berger[1], B. Garzón[2], C. Driver[1], L. Helbling[1], M. Tomasik[1]

[1]Institute for Educational Evaluation, Switzerland
[2]Institute of Education, University of Zurich, Switzerland

The COVID-19 pandemic accelerated the digitalization of assessments. However, an increased number of test results does not automatically provide better feedback. We present an approach for linking the results from a set of standardized high-stake assessments (HSA) to those of an online item bank for formative assessment (IFA). Four compulsory standardized high stakes assessments provide summative feedback on the students' ability at four important time points during the students' school career. Between those measurement occasions, teachers and students can use an IFA for data-based decision making. Both assessment tools have been used over several years by thousands of students. Data from the HSA were scaled using a separate 2PL IRT model for each grade and each cohort. Data from the IFA was scaled concurrently using one single 2PL IRT model. For the main analyses, we correlated achievement data from those IFA that were closest in time to the HSA under investigation. Overall, the results suggest a moderate predictive value of the IFA score for the HSA result. We discuss the implications of these findings for the potential of predicting HSA results from data collected in an IFA situation and we point to future methodological possibilities of these linking approach.

Going gradeless. A scoping review of reduced grading in assessment

D. Normann[1], L.V. Sandvik[2], H. Fjørtoft[3, 4]

[1]Norwegian University of Science and Technology (NTNU), Norway
[2]NTNU, Norway
[3]NTNU Norwegian University of Science and Technology, Norway
[4]Nord University, Norway

There is considerable interest in changing grading practices to ameliorate their perceived negative impacts on student learning and motivation. Therefore, reducing the role of grades in assessment is becoming popular in various contexts. We conducted a scoping review of 21 peer-reviewed studies that explored the reduced use of grades. Using qualitative content analysis and a theory of action framework to investigate the rationales for going gradeless, we analyzed data to identify changes in assessment practices and to explore how such changes were perceived by teachers and/or students. Six key factors were identified: contextual conditions (e.g., policy changes afforded reduced grading), rationales (e.g., improving psychological or socio-relational dimensions), changes in assessment practices (e.g., from grades to pass/fail), conditions for successful implementation (e.g., challenging assessment beliefs), impact on student learning (e.g., improving well-being or impacting student attitudes), and stakeholder concerns (e.g., negatively impacting study habits). Furthermore, although teachers perceived benefits of going gradeless, students' views were divided. We suggest that building a theory of action approach could help support practitioners in understanding contextual needs, exploring existing assumptions about grading, developing an appropriate action strategy, and evaluating the outcomes of going gradeless.

Discussion Group 1

11:00 - 12:00

‘Measurement’ or ‘assessment’? What does language, and how we use it, tell us about our understandings of assessment?

I. Nisbet[1], S. Shaw[1], L. Wiseman[2]

[1]University of Cambridge, United Kingdom
[2]University of Glasgow, United Kingdom

This Discussion Group will examine the language we use, in different contexts, to talk about assessment - and particularly the concept of “measurement”, as used to describe assessment. The stimulus for that discussion will be three assessment scenarios to which Discussion Group participants will have access in advance, each relating to the same student: two in education (in the broadest sense) and one in health. Participants will consider the nature of each assessment and the language used to describe it and in communication about it within and beyond particular disciplines. This should lead participants to reflect on the language they use and what it means to them, and on its relationship to the contexts and cultures within which they work and describe their work to others. Such reflection should promote discussion of the extent to which we share understanding of the language we use to talk about assessment.

Discussion Group 2

11:00 - 12:00

Crossing the line: Should technology drive and inform what we teach, learn or assess?

R. Hamer[1], G. Hudson[2], D. Seabrook[3], S. Shaw[4]

[1]International Baccalaureate, Netherlands
[2]GA Partnership Limited, United Kingdom
[3]Qualifications Wales, United Kingdom
[4]University of Cambridge, United Kingdom

A core principle of test design is creating tests that allow those tested to demonstrate their knowledge, skills and understanding in ways that can be evaluated fairly and reliably. When times and circumstances change, familiar methods of assessing require a rethink and new approaches and tools may need to be considered. Such change may offer new opportunities but often require new knowledge and skills to work well. Changes introduced through technology might be enriching, resulting in more authentic assessment, or conversely may reduce the construct validity and fairness of what is assessed.

The past two years have seen major shifts in education, with learning and teaching as well as exams and assessments moving digital. We all needed to adjust, learn to use technology at speed and in ways not imagined before. Consequently, there is a wealth of experience of benefits and drawbacks, all feeding into personal convictions and perspectives on the value and efficacy of using technology and how it impacted the learning, teaching and assessment. This Discussion Forum aims to access these experiences and perspectives and unique solutions found, by actively engaging all participants in debates focusing on the degree to which technology should drive the learners’ educational experience.

Discussion Group 3

11:00 - 12:00

Going digital: Developing and administering digital numeracy assessments for primary schools

G.A. Nortvedt[1], K.B. Bratting[2], A. Pettersen[3], O. Kovpanets[3], H.H. Haram[1]

[1]University of Oslo, Norway
[2]Universitetet i Oslo, Norway
[3]UiO, Norway

In early primary school grades, teachers often use assessment tools to identify students at risk of falling behind in mathematics or numeracy. Traditionally, these assessments have been paper based, and students have had teacher support when completing the assessments. More recently, increasing numbers of assessments are being digitalised. Digital assessments may include, as one example, interactive task formats that replace the static paper-based formats. The aim of this discussion group is to discuss the potential advantages of digitally formatted assessments and confer on issues related to developing and administering digital numeracy assessments for young students (5–9-years-old). For example, the digital format may provide visual and linguistic support for young students who do not yet read well, but at the same time, these students' motor skills, experience with digital assessments and previous exposure to game-based apps may pose challenges to test development and test taking. Identifying and addressing these issues is vitally important to ensuring effective digital assessment practices as we continue to move deeper into the digital domain

Discussion Group 4

11:00 - 12:00

Studying, researching and networking during a pandemic: issues faced by post-graduate and early career researchers and how a learning, collaborative community can help

S. Manassian[1], D. Normann[2], J. Leonardsen[3]

[1]PSI Services UK Ltd, United Kingdom
[2]norwegian university of science and technology (NTNU), Norway
[3]NTNU, Norway

The aim of the Post Graduate Student and Early Researcher Network is to create a supportive and caring environment where students and researchers can find a community of people experiencing similar challenges. This need underpins the discussion session that the steering group would like to propose. The Network would like to invite all current members and any interested participants at the conference to join us at a discussion session where we will explore the conference theme in relation to our experiences as students and as assessment researchers.

Over the last two years, many of us have experienced feelings of isolation and have had to deal with issues of access to resources and people in order to progress our studies and/or work. This discussion session will allow us to explore a series of four questions and to share our experiences of studying and working in uncertain times. The key areas for discussion are: the challenges and opportunities in assessment, how changes in assessment have impacted the institutions where we work and study, how we can actualise new visions for assessment and how the Network can support its member.

Discussion Group 5

11:00 - 12:00

Engaging teachers in professional development in today's uncertain times

M. Talbot[1], M. Walker[1]

[1]University of Leeds, United Kingdom

The discussion will augment work planned for summer 2022 at the University of Leeds to discover how best to support teachers to engage with professional development (PD). PD provision has been much-affected in recent years by the diminished role of Local Authorities in providing training, the rise of Multi-Academy Trusts offering trust-specific opportunities, and the Covid-19 pandemic, so we are keen to get an up-to-date picture. We would like to include not just teachers but school leaders, teacher trainers, and other educational professionals in the discussion.

We want to explore experiences of teachers' enhanced role in assessment in 2020 and 2021 and to consider how this role might continue to change in these uncertain times, including what should be assessed, by whom, and how. Appropriate PD could help teachers address the current gap between the ways students are taught and learn and how they are assessed. This is linked to curriculum reform, teacher capacity, and initial teacher education, and would require time and money to embed changes in their practice. We suggest PD related to educational assessment forms the next logical step and represents an "evolution in policy and debate about the broader role of assessment in education".

Assessment Cultures III

14:00 - 14:30

Evidencing Transversal Competencies: Student Centred Approaches to Developing Transcripts for Creativity and Curiosity

S. Krstic[1], S. Richardson[1]

[1]ACER, United Kingdom

This paper presents the findings from the work ACER did to define conceptual frameworks for creativity and curiosity and to develop transcripts for recording student achievement. Our goal is to enable learners to evidence their creativity and curiosity, and teachers to support and recognise it. The setting for this work are the four International Baccalaureate programmes, delivered in 155 countries.

The frameworks draw on extensive scholarly literature to define the tangible skills, knowledge and attributes that are representative of creativity and curiosity in various classrooms. They are accompanied by a learner profile mastery transcript in which learner achievement can be recorded, and progress tracked. Uniquely, we have used the concept of a thermometer to help students and teachers identify factors that cause curiosity and creativity to rise and fall, rather than a standard linear progression.

Ultimately, the intention is that schools and teachers will be able to encourage, recognise and support learners to strengthen their 21st century skills, better preparing them for their future lives. It is hoped that the findings from this research will have implications and value for all schools. Moreover, we believe that it establishes a learner-centred approach for the evidencing of other 21st century skills.

Re-conceptualising student assessment literacies: promoting critical participation as students embrace their roles in assessment.

C. F. Correia[1]

[1]University College London - IoE Faculty of Education and Society, United Kingdom

Assessment literacy has been advocated as a key for stakeholders' active participation and critical engagement with educational assessment. Several principles and models for the development of teacher assessment literacy have been identified. However, less attention has been paid to students' assessment literacy. Advances in theory and practice seek to promote students' voice and active participation in assessment. However, the reality is that 'assessment is still something that is done to students rather than with students', and across all educational levels.

Assessment literacy can be seen from different theoretical lenses. This paper will explore students' empowerment in assessment, by re-conceptualising student assessment literacy focusing on the multiple roles of the student in assessment. The paper is guided by the following questions: 1) What are the characteristics of student critical participation in feedback and self-assessment in terms of knowledge, beliefs, attitudes, and practices? 2)What kinds of assessment literacies are needed to support critical participation in assessment? .These are addressed using sociocultural theory and participatory frameworks to explore student engagement with feedback and self-assessment. The paper seeks to critically review the literature to explore factors that support and hinder in the adoption of these roles, from a student assessment literacy perspective.

Different Consequences and Contexts Produce different effort on ILSAs

A. Zhao[1, 2], G. Brown[1], K. Meissel[1]

[1]The University of Auckland, New Zealand
[2]The Ministry of Education, New Zealand

International large-scale assessments (ILSA) report between-country performances assuming that greater performance is explained by greater ability rather than motivation. However, because students' test-taking motivation does predict performance, if the importance of ILSAs differs across jurisdictions, it is possible that country reputation consequences may elicit different effort in different societies. This study compared student self-reported effort and test-taking motivation in Shanghai and New Zealand, jurisdictions that range from high to middle on PISA. A between-subjects experiment systematically assigned senior secondary school students in Shanghai (n=1003) and New Zealand (n=479) to one of the three consequence vignettes (i.e., none, my country, or me personally).. Students completed the Student Conceptions of Assessment (SCoA) inventory before being assigned to the condition in which they described their effort and motivation. Factor analytic methods established structures and comparability, while structural equation modeling examined impact of conceptions of tests to test-taking motivation. Latent mean analyses indicated that students' reported effort, anxiety, and importance were lowest in both contexts when no consequences were attached. However, the gap between personal and country stakes was much larger in New Zealand, suggesting that PISA results may reflect differences in cultural norms around the importance of trying on country related ILSAs.

Psychometrics & Test Develpoment III

14:00 - 14:30

Critical thinking and logical thinking assessment in modern education: computerized performance-based assessment approach

I. Uglanova[1]

[1]National Research University Higher School of Economics, Russia

Critical thinking and logical reasoning are important predictors of academic achievement and key educational objectives in modern education. In the research, we present a new assessment tool with automatic scoring in the format of a computerized performance-based assessment (CPBA). The aim of the research is to examine the psychometric properties of the CPBA and to investigate the patterns of critical and logical thinking within the tasks. The conceptual framework was developed based on Jean Piaget's theory (combinatorial logic factor) and modern studies on critical thinking in psychology and education (analysis of information and making inferences factors). The sample consists of 5294 students aged 10 to 15 years (5-8 grades). The analysis was based on the methodology of Cognitive Diagnosis Modeling. The results revealed that the theoretically expected three-factor structure was confirmed. Three classes were filled with the majority of students: the possession of none of the factors, all of the factors, and all of the factors but not the combinatorial logic factor. The results are in accordance with the theoretical expectation since combinatorial logic is a characteristic of the most mature stage of logical thinking (the formal operational stage).

A consideration of factors affecting the use of Automatic Item Generation (AIG) in developing items for use in high-stakes assessments.

G. Cherry[1], C. Scully[1], M. O'Leary[1]

[1]Dublin City University, Ireland

Test items are essential for assessing learners' knowledge and/or understanding. However, constructing items using traditional, manual methods can be a challenging, expensive, and time-consuming process that involves many trained experts including item writers, subject matter experts (SMEs) and psychometricians. Automatic item generation (AIG) has emerged in recent decades as a response to the challenges involved with traditional item development. AIG is a rapidly evolving research area where various methods are used to generate a high volume of items using computer technology and item models based on a set of rules (algorithms). As the majority of items used in commercial testing continue to be written by humans, AIG may prove to be a valuable resource in terms of enhancing and shortening the overall item development process. The purpose of this paper is to detail the applicability and relevance of AIG for developing items for use in large-scale, high-stakes assessments. Key findings focusing on the benefits of AIG and critiques arising from the literature are presented. In addition, a series of recommendations are provided to aid decision making regarding incorporating AIG into the item development process. The paper concludes by highlighting several areas that are deserving of future research in this area.

How much is enough? Reducing assessment time and maintaining reliability

S. Hughes[1], R. Clesham[1]

[1]Pearson, United Kingdom

Assessment organisations have a responsibility to ensure that they are not assessing learners more than is necessary. Lengthy tests are burdensome for learners and may present threats to validity if the testing time leads to fatigue, distraction, or low motivation. Assessment organisations also face a burden in producing test content and providing scoring services for lengthy tests. If a similar level of reliability and validity can be achieved in a shorter testing time, then it is in the interest of testing organisations and learners alike to shorten the test. This presentation summarises the research efforts that supported reducing the testing time of a large-scale high-stakes computerised language assessment by one third. This presentation will be of relevance to researchers and practitioners interested in creating effective and efficient assessments.

Assessment Against the Backdrop of Covid IV

14:00 - 14:30

A-level students’ experiences of practical science at home and school during the Covid-19 pandemic and the impact on transitioning to university

K. Finch[1], C. Balaban[1], H. Cramman[2]

[1]AQA, United Kingdom
[2]University of Durham, United Kingdom

A collaborative project between AQA, Durham University and the University of Liverpool explored A-level students’ experiences of practical science during the Covid-19 pandemic and the implications for transitioning to university.
Online focus groups were conducted with A-level students from eight centres in England. First-year undergraduate science students were then invited to participate in an online survey. Four main themes emerged from the focus group data: differing student experiences of practical science, with variance at a centre, subject and teacher level; a gap in manipulative skills; improved scientific enquiry and soft skills through learning at home; a positive attitude towards transitioning to university. Survey data showed that almost all students were able to experience practical work during their A-level studies. This involved a range of practical skills, and for some students it also included designing their own procedures and experiments. Students reported that, in general, they felt well prepared for higher education, although the findings indicate that they might need more support to work independently in laboratories.
This research highlights that students had varying experiences of practical science during the pandemic. The findings can be used by higher education institutions to help them ensure a smoother transition for their incoming students.

14:30 - 15:00

Estimation of component marks during a pandemic

C.E. De Groot[1]

[1]Cambridge Assessment, United Kingdom

The Covid-19 pandemic has shown that pandemics can create significant disruption to periods of assessment. In “normal” times, awarding organizations use well-established methods to impute marks for students that miss an exam component due to illness. However, in “pandemic” times, having large numbers of students missing an exam component may call these methods for imputing marks into question. This paper uses Monte Carlo simulation techniques to study the robustness of imputation methods when large numbers of students miss a component exam. While standard methods appear relatively robust, we find that they are prone to bias when a disproportionate number of high (or low) achieving students miss a component. We propose an adjusted-Z-score method to correct for this potential bias.

Fairness & Social Justice II

14:00 - 14:30

Path to Inclusion and Accessibility: Improving the Access and Use of Modified Exam Papers to Level up Opportunities for Students with SEND

L. Liu[1], K. Mason[1], B. Redmond[1], H. Dalton[2], G. Grima[3]

[1]Pearson, United Kingdom
[2]Pearson Education, United Kingdom
[3]Pearson UK, United Kingdom

Modified exam papers provide reasonable adjustments, such as enlarged papers, fonts, interactive pdf, and video clips for listening, to support students with Special Education Needs and Disabilities (SEND) to perform high-stakes exams. However, the research into the access and use of modified papers has been significantly underexplored. To advance understanding of why schools and students use different types of modified papers and how these modifications support or hinder students from accessing exam papers, we surveyed exam officers, students aged 13-16 across the UK, and followed up with student interviews. The findings show that students favour some modifications, including enlarged papers, fonts, diagrams, and listening functions. These well-designed features effectively reduce their reliance on other human readers and significantly boost their confidence in taking exams solely. Nevertheless, over a third of the exam officers and students reported the modifications, such as interactive pdf, braille papers, etc. have never or less likely been applied in their daily classroom settings. The discrepancies between classroom learning and exam-taking may lead this group of students even more disadvantaged.

14:30 - 15:00

Evaluating sources of differential item functioning in high-stakes assessments in England

Y. El Masri[1], Q. He[1]

[1]Office of Qualifications and Examinations Regulation, United Kingdom

The Covid-19 pandemic has placed considerable pressure on educational assessment systems. It has highlighted social inequalities, including those that are reflected in students' performance in high-stakes examinations. In its continuous efforts to ensure that regulated qualifications are valid, reliable and fair, and in its commitment to equality, diversity and inclusion (EDI) principles, the Office of Qualifications and Examinations Regulation (Ofqual) in England is undertaking research to evaluate the extent to which a mixed-method approach can be used to identify items that may disadvantage specific subgroups of candidates. In doing so, Ofqual hopes to foster the development of more accessible and fairer assessments.

The research comprises two phases. The first phase consists of a differential item functioning (DIF) analysis based on Rasch modelling and a weighted item mean score (WMS) approach and examines item behaviour in assessments from a variety of high-stakes qualifications regulated by Ofqual, with respect to different subgroups. In the second phase, subject experts evaluate selected DIF and non-DIF items with consideration given to item features that could potentially affect the performance of specific subgroups.

Results will be presented with examples of items to illustrate potential sources of DIF. The usefulness of the mixed-methodology approach will be discussed.

15:00 - 15:30

Defining Inclusive Assessment: Can assessment meet the needs of all students?

N. Care[1]

[1]Assessment MicroAnalytics, United Kingdom

Inclusion is closely tied to fairness in assessment, for an assessment cannot be fair if candidates are discriminated against due to construct irrelevant factors, such as race, gender or disability. However, whereas the application and meaning of fairness for educational assessment has been explored considerably, the same cannot be said for inclusive assessment. In the case of inclusive assessment there is no established definition, resulting in a discipline that cannot assess how and when inclusion is achieved in the creation of educational assessment. In this paper we consider the complexities of defining inclusive assessment whilst offering a definition and theoretical framework for the use of test-designers and assessors.

This paper presents an argument and evidence-based approach to inclusion. By accepting and understanding the limitations of assessment as a valid and legitimate attempt to capture a person’s ability, we argue that it is possible to build a framework for inclusive assessment that balances the needs of heterogenous populations with the limitations of assessment. The hope being that with clear definitions and frameworks there will come a stronger impetus for the creation of inclusive tests and improved access for candidates with disabilities and from diverse backgrounds.

Comparative Judgement II

14:00 - 14:30

Unpacking decision making in comparative judgement: A stimulated think-aloud methodology to gain insight into young peoples' decision making

E. Hartell[1, 2], J. Buckley[3, 4]

[1]KTH Royal Institute of Technology, Sweden

[2]Haninge Municipality, Sweden

[3]Faculty of Engineering and Informatics, Technological University of the Shannon: Midlands Midwest, Athlone, Ireland, Ireland

[4]Department of Learning, KTH Royal Institute of Technology, Sweden

While reliability has been the subject of much Comparative Judgement (CJ) research, understanding its validity, which and relates directly to the included judges and their decision-making, is paramount. Understanding this decision-making better would add significantly to the formative use of CJ and its use in educational task-design.

This paper reports on a pilot study exploring a novel methodology aiming to unpack judges' decision-making. One 11-year-old student completed a CJ session on a selection of portfolios developed in response to an authentic design-task in STEM education. During this, a novel "stimulated think aloud protocol" was implemented, which was developed by synthesising aspects of traditional think-aloud-protocols with stimulated recall interviews.

The approach is considered to have worked well as it was immediately evident that prompts were required to keep the participant on task and to continue verbalising their thoughts. As in this case the participant was younger, giving support in what to verbalise appeared necessary. The approach was possibly more useful due to the age of the participant. Limitations exist in that the stimulating prompts could influence participant decision-making if they provoke reflection which otherwise would not have occurred in an undisrupted CJ judging session.

14:30 - 15:00

Assessment of Art and Design Courses using Comparative Judgment in Mexico and England

K. Mason[1], L. Garelli[2]

[1]Pearson, United Kingdom
[2]Anahuac University, Mexico

This paper reflects on two experiments using pairwise comparative judgements to assess art and design courses. This method has many potential advantages over traditional rubric assessments, such as the ability to cope with the wide styles and types of assessment evidence produced in these courses. The first describes how comparative judgement was used in an undergraduate painting course in Mexico, and the second applied the method to an art and design course for 16-19 year olds in England. The studies showed high degrees of reliability in the scales created from the judgements, and most judges displayed acceptable levels of infit, suggesting a good level of agreement about the qualities required for a good piece of work. Participants also commented that ensuring each piece of work was judged by a range of judges may also contribute to an increased fairness in the assessment. However, there are many practical challenges still to overcome before this method could be operationalised in a national assessment.

How do judges in Comparative Judgement exercises make their judgements?

T. Leech[1], L. Chambers[1]

[1]Cambridge University Press & Assessment, United Kingdom

The use of comparative judgement (CJ) is increasing in various assessment contexts. Simultaneously, emerging developments in technology mean assessment processes are progressively moving to being digitally facilitated – something partially fuelled by the Covid-19 pandemic. Online CJ is being considered as an alternative for existing assessment processes including standard maintaining, and is seen to be reliable and practical. However, we consider the under-explored validity angle, discussing what processes judges use to make their decisions and what features they focus on when making their decisions. We report the results of both a study into the processes used by judges, and the outcomes of surveys of judges who have used CJ. We develop a four-dimension model to explore what features have an impact on what judges attend to and explore the distinctive ways in which the structure of the question paper, different elements of candidate responses, judges' own preferences and the CJ task itself affect decision-making. We conclude by questioning, in light of these factors, whether the judgements made in standard maintaining, whether using CJ or not, are meaningfully holistic.

Educational Policy and Assessment III

14:00 - 14:30

Assessing civic and citizenship education – recurrent challenges and open issues in the Italian context

V. Damiani[1], G. Agrusti[1]

[1]LUMSA University, Italy

Civic and citizenship education (CCE) is an educational area that lingers in the national curricula although it has become over the years a key theme within the Italian education system and in 2019 civic education was reintroduced as a specific subject. Its promotion lacks implementation conditions, monitoring actions, and data.

Research conducted internationally on CCE represents unique sources for the assessment of CCE in Italy and comparatively in relation to other countries.

This paper presents an analysis of civic and citizenship education by integrating the most relevant data from the latest Eurydice report (2017) with the results of the second cycle of the IEA-International Civic and Citizenship Education Study (ICCS, 2016).

It analyses some elements that characterize civic and citizenship education in Italy in comparison with other European countries. Five areas are examined, namely: CCE approaches; objectives and key themes; participation of teachers, students and parents at school; assessment; teacher training (initial and in-service).

The final section highlights open issues and challenges in the Italian context for CCE, mainly related to the misalignment between instructional design and assessment and the lack of teacher training and participation opportunities at the school level for students and parents.

The future of qualifications and assessment in England: Exploring a coherent curriculum framework for numeracy and literacy.

L. Watts[1], H. Dalton[2]

[1]Pearson, United Kingdom

[2]Pearson Education, United Kingdom

Drawing on large-scale, mixed-methods research taking place in 2021 and 2022, this paper explores how clearer coherence between the aims of curriculum and assessment could be transformational for young people in England. With curriculum, qualifications and assessment inextricably linked, this research explores that without policy coherence threading through the various elements, the ‘problems’ education may be designed to ‘solve’ (smooth transition to the labour market, the ability to thrive in adulthood) risk being misaligned.

One such area of contemplation is the provision of numeracy and literacy skills in the 14-19 phase. The Department of Education in England recently set out ambitions to drive up adult literacy and numeracy standards and average GCSE grades in Mathematics and English. Using an exploratory approach, we unpick the current state of (in)coherence in the curriculum across the lower and upper secondary stages, drawing on international approaches to a more fundamental intention to enable all students to become numerate and literate into adulthood.

Assessment and the Vision for Learning: Synergies between the Draft Primary Curriculum Framework and the Framework for Junior Cycle

S. Tuohy[1], T. Curran[1]

[1]National Council for Curriculum and Assessment, Ireland

Changes to how assessment is positioned and used have been central to recent education reform in Ireland. This paper explores key messages espoused in two curriculum policy documents: the Draft Primary Curriculum Framework and the Framework for Junior Cycle, demonstrating how these assessment reforms reflect the priorities of 21st century education in Ireland.

The Draft Primary Curriculum Framework has undergone extensive consultation over the past two years and provides a vision for children’s learning across the eight years of primary school emphasising that assessment is central to teaching and a fundamental aspect of teachers' daily practice. The Framework for Junior Cycle led to intense debate and discussion about the purpose of assessment in lower secondary education and places significant focus on assessment supporting learning. It sets out a dual approach to assessment including the introduction of Classroom-Based Assessments and social moderation during Subject Learning and Assessment Meetings.

Three key inter-related tenets common across both Frameworks are discussed: assessment as an integral part of teaching and learning; assessment can serve multiple purposes and assessment as a learner-centred and collaborative process. Each of these is explored in detail with a focus on how learners, teachers and schools are empowered across both Frameworks.

Perspectives of End-users and the General Public on Assessment I

14:00 - 14:30

Indignation, toxic narratives, and qualification (re)design

P. Newton[1]

[1]Ofqual, United Kingdom

In this presentation, we will explore the hypothesis that certain threats to validity – concerning certain varieties of assessment error – are less tolerable than others. In particular, where assessment inaccuracy appears to reflect partiality, prejudice, or blatant malpractice, this tends to ignite high levels of indignation among members of the public, which identifies error of this sort as especially intolerable. Importantly, other stakeholders, including curriculum and measurement experts, do not always see things similarly. Three case studies relating to the use of teacher assessment in General Certificate of Secondary Education (GCSE) qualifications in England will be used to explore this hypothesis. In each case, we can see the emergence of a 'toxic narrative' that crystalised a strong sense of indignation felt by many. This helped to construct a tipping point in public opinion, from tolerance of the teacher assessment approach in question to intolerance. Following this pattern, it appeared that: the 'cheating students' narrative sealed the fate of GCSE coursework during the 2000s; the 'cheating schools' narrative sealed the fate of GCSE controlled assessment during the 2010s; and the 'algorithm' narrative sealed the fate of the GCSE teacher assessment moderation model in 2020.

14:30 - 15:00

Research claims within the Education Industry (EI): Managing reflective practice

S. Shaw[1], S. Fitzsimons[2]

[1]University of Cambridge, United Kingdom
[2]International School of Brussels, Belgium

The Education Industry (EI) is a far reaching, innovative and rapidly evolving field of business. To ensure success and integrity in the EI, organisations and companies strive to deliver high quality products and services in an efficient and ethical manner. Education research plays an important part in the EI by underpinning product and service developments and through illustrating impact. Organisations and companies also share these research claims when marketing to potential customers and investors. However, there can sometimes exist a disjunction between those conducting research and those responsible for interpreting the research for the purpose of public dissemination. This presentation seeks, firstly, to investigate what constitutes an education research claim (particularly in relation to educational assessment claims). The risks associated with such claims are then identified and a review process suggested so educational bodies can ensure accuracy and ethicality in their claims. Adopting a generalised case study approach, educational claim-making is contextualised from the stance and perspective of an awarding organisation.

15:00 - 15:30

Continuity and Change: a case study of developing next generation vocational qualifications

E. Carey[1], E. Boyd[1]

[1]Scottish Qualification Authority, United Kingdom

SQA is engaged in a major programme of work to fundamentally reshape our flagship Higher National Qualifications that will have meta-skills development at their core. This session will outline key aspects of SQA's research activity and work with stakeholders in Scotland that shapes this work.

Higher National Qualifications are vocational qualifications designed to meet the needs of employers, by developing employees with the technical and interpersonal skills needed to support industry. The structure of the current qualifications was last revised in 2002 and was completed in 2008. Since then, SQA has adopted a cycle of review to keep products up to date, using Qualification Support Teams to ensure the products continue to meet customer need.

Much has changed in the Educational Policy and Assessment environment since then, with numerous new government strategies in place and significant changes in both industry and society, driven by the digital revolution and now set against the backdrop of the COVID-19 pandemic.

SQA, using Next Generation Higher National as a case study, wish to discuss some of the challenges of developing vocational qualifications in smaller nation, while meeting the needs of employers and learners in a time of shifting paradigms in the world of work.

Ignite Presentation Session

16:00 - 16:15

Providing an evidence-base to inform digital assessment design

S. Hughes[1]

[1]Cambridge University Press and Assessment, United Kingdom

A joint programme of work at OCR and Cambridge Assessment International Education is developing digital assessments informed by research. This presentation focuses on the research process in a digital context and covers three areas.
1) The programme of research.
This includes three strands with different purposes, to:
• inform decision making during assessment design
• build in assessment quality from the start
• address big questions about the digital assessment context
2) How assessment development working practices impact on the research process.
Using design thinking and agile working to develop assessments means that researchers need to keep close to research users to define questions, share interim findings and adapt to their needs. The research process needs to be faster than it is traditionally, and reporting takes on different forms.
3) How the research has impacted assessment design decisions.
For example, in relation to what constructs should be assessed and how.

16:00 - 16:15

Cultural Challenges In Developing An Assessment For Indian Children During A Pandemic

C. Jellis[1]

[1]CEM, Cambridge University Press and Assessment, United Kingdom

Presenter, Dr Chris Jellis, Senior Research Associate
Conference subthemes: Test Development, Assessment Cultures, International Assessments E-assessment, Psychometrics and Test Development

The pandemic has led to remote schooling and working around the world, forcing assessment developers to work in new ways. In ‘normal’ times, cross-cultural challenges in development are often uncovered, explored, and addressed through observation and face-to-face discussion.

In this talk, I outline the process of creating a computer-based assessment for 4 to 5-year-olds in India. The development team worked from home, and I evaluate five key cultural challenges that were identified and resolved remotely. These were around: initial assessment evaluations, creating new content, maintaining an appropriate cultural balance, and trialling during lockdown.

Ultimately, I aim to show that it is possible to develop a new, culturally appropriate assessment under less than ‘normal’ circumstances.

Narrated histology videos as a medium of streamlining teaching in a medical curriculum. Does student satisfaction reflect student performance?

D. McLaughlin[1], K. Clarkson-Dornan[1], C. Foy[1]

[1]Queen's University Belfast, United Kingdom

Introduction: Histology is a crucial element of medical education, enabling students to understand normal organ systems. Task-based activities, including ‘identify and justify’, are used to keep students focused on topic. The goal is to encourage students to ‘learn by doing’. Methods: With full ethical approval, voluntary participants were recruited and given access to narrated histology videos between pre- and post-tests. A survey gathered student opinion on use of videos.
Results: Qualitative feedback was predominantly positive with students strongly agreeing that videos could be used as a revision tool. Students felt that videos enhanced knowledge of the subject; a result that reached statistical significance. Quantitative analysis demonstrated a statistically significant, negative correlation for control and experimental groups when assessing pre- and post-test scores.
Conclusion: Student opinion and satisfaction provides valuable information into engagement with educational resources. However, data suggests that self-perceived ideas of knowledge enhancement were not reflective of assessment scores.

16:00 - 16:15

Data forensics or how to watch the assessment in the times of pandemic

N. Curkovic[1], J. Bugarin[1, 2], S. Fulgosi[1]

[1]National Centre for External Evaluation of Education, Croatia

[2]Faculty of humanities and social sciences, University of Zagreb, Croatia

Cheating on exams is an issue from different perspectives. It causes a decrease in measurement accuracy and consequentially affects the measurement validity. When focusing on test-result use, especially on high-stakes exams, cheating puts honest examinees at a disadvantage. There are different strategies for cheating prevention. One of them is to have many persons overseeing test takers. However, pandemic conditions emphasized reducing the number of people involved in the test delivery process. In the Croatian context, this meant reducing the number of human proctors. Hence, this research aimed to apply different data forensics procedures to examine and compare the usage of statistical indicators of cheating. Research has shown that statistical procedures could be very effective in detecting potential cheating. Omega index turned up to be the most appropriate statistic that could also be used in planning measures for cheating prevention.

Expanding your horizons: How adding flexibility to your authoring is the key to unlocking new visions for assessment

S. Crowley[1]

[1]GradeMaker, United Kingdom

The pandemic has prompted many assessment organisations to review their exam process and delivery models. This presentation proposes that the path to modernisation starts with future-proofing your exams authoring process, applied as part of a 'modular' technical infrastructure.

The presenter will explain why adopting authoring technology in this way provides the best platform for adaptation and innovation, offering awarding organisations the flexibility to:

- Transition from print to onscreen testing as gradually or quickly as required
- Choose 'best of breed' technology components such as adaptive engines, test players, marking and e-proctoring systems – and swap components over time without affecting content creation
- Re-use and re-package legacy test items, for example, to provide formative testing services
- Support a future move to 'on demand' testing by transitioning to an item banking authoring model
- Address workflow inefficiencies and security audit requirements by automating parts of the authoring process

16:00 - 16:15

What can we learn about the validity of questionnaire responses by using response process data?

H. Eklöf[1], E. Lundgren[1]

[1]Umeå University, Sweden

Although computer-delivered tests are not new on the assessment scene, they have become more common in recent years and the future vision for assessment is likely even more digital. With digital assessments come new possibilities to explore assessment validity and student behavior through traces students leave on the computer during assessment. The current study emanates from previous work on developing a response process modeling approach for assessing student latent test-taking effort on the PISA test through their behavior on the student questionnaire. As findings from this work suggested odd properties in some of the items, the purpose of the present study is to use the same modeling approach and discuss, within the framework of satisficing theory, validity issues in questionnaires in general, and specifically in the PISA context. If some item types are more affected by aberrant response behavior, this could have consequences for the validity of interpretation of findings.

An exploration of the use of online clinical examinations during the pandemic- what have we learned?

A. Patterson[1], M. Hennessey[1], L. Courtney[1], E. Burke[1]

[1]Trinity College Dublin, Ireland

During the COVID crisis accessibility to patients and infection control issues caused problems for assessment in healthcare programmes, both at undergraduate and postgraduate levels. Arising out of pragmatism, many schools transferred their clinical examinations to an online format in response to these limitations. Virtual OSCES and online clinical examinations were created out of necessity to ensure the progression and graduation of students required to enter the healthcare workforce. This presentation will look at the literature published since 2020 on the use of online clinical examinations and modified in person examinations in a range of healthcare professions, the variation of approaches used, the reliability and validity of the data reported and the acceptability of these changes to a range of stakeholders.

16:00 - 16:15

Collaborating to successfully deliver quality e-assessment practice

E. MacIntosh[1], G. Clark[1]

[1]SQA, United Kingdom

The COVID-19 pandemic has been a great disruptor in education, shifting learning and teaching online as we go through lockdowns and forcing traditional assessment models to adapt to a digital space.

As Scotland's awarding body, providing a comprehensive suite of National Qualifications as well as vocational qualifications, this has provided SQA with both challenges and opportunities. What has become apparent is that in order to best support our centres, many of whom are new to the area of e-assessment, we need to work with them to understand the picture on the ground. In order to do this more effectively, we have been working collaboratively on a programme of mixed-method research. Work with learners and practitioners across schools, colleges and training providers has given us both a high-level and a more in-depth understanding of how e-assessment has been used across our centres, and what expectations are for the future.

16:00 - 16:15

Conceptualising the Impact of Germane Cognitive Load on the Assessment of Higher-Order Thinking Skills

G. Sultanova[1]

[1]NIS, Kazakhstan

To successfully compete in the labour market, especially in times of uncertainty resulting in a sharp rise of skills mismatch, the workforce have to possess twenty-first century competencies such as problem solving, critical thinking, and creativity. Hence, education at all levels should produce graduates possessing these competencies that are based on higher order thinking skills (HOTS), namely the abilities to apply, analyze, evaluate, and create knowledge. In the case of HOTS, the cognitive load of tasks is crucial. Thus, there is a need to examine the theoretical relationship between a task's cognitive load and the assessment of thinking skills required for the task accomplishment. The aim of this study is to find out how to improve assessment procedures so that they can contribute to better performance of complex tasks in secondary education. For this purpose, Bloom's Taxonomy of Educational Objectives and Cognitive Load Theory are interrelated and operationalised. The research questions and hypotheses are outlined. In this empirical study, the path analytic method is applied to estimate the magnitude and strength of effects within a hypothesized causal system. This method allows predicting the impact of a task's cognitive load on the assessment of thinking skills required for the task accomplishment.

Using feedback to enhance both academic skills and explicit wellbeing skills in schools: presentation of a novel evidence-based tool.

E. Lucciarini[1, 2], N. Bressoud[3, 4], P. GAY[5]

[1]Valais University of Teacher Education, Switzerland
[2]University of East London, United Kingdom
[3]Valais university of teacher education, Switzerland
[4]chEEERS lab, University of Fribourg, Switzerland
[5]Vaud University Teacher Education, Switzerland

Feedback is known to be one of the most effective ways to enhance pupils' academic skills (Hattie, 2018). However, its implementation remains challenging for teachers (Carless & Boud, 2018; Dawson et al., 2019). Thus, this intervention aims to present the feedback ruler: a novel evidence-based model and tool designed to fit into teaching practices. The feedback ruler is built on a literature review (Lucciarini, 2020) and the self-determination theory (Deci & Ryan, 2012). In addition, since feedback is primarily used to boost academic skills (e.g., Hattie & Timperley, 2007), and that there is a literature gap around its effects related to wellbeing skills, the feedback ruler also addresses the latter issue.
In short, the ruler has two scopes: 1) meeting teachers' needs by offering an easy-to-use and evidence-based feedback model; 2) developing both academic and wellbeing skills through feedback. During this presentation, we will present the tool and its theoretical background. Questions and feedback from the audience should permit the authors to enhance the tool and its implementation in school contexts.

Exploring the role of Assessment Literacy in times of uncertainty

16:00 - 16:15

Teachers’ Language Assessment Literacy during COVID-19: What have we learnt?

D. Tsagari[1]

[1]Oslo Metropolitan University, Norway

With the global outbreak of COVID-19, educators worldwide have encountered challenges in implementing planned on-site assessment. Most, if not all countries had to announce responsive measures. The pandemic prompted debates on the shift to Emergency Remote Teaching (ERT) and ways of coping with the New Normal. This study aimed to explore assessment measures and practices in different countries. It specifically investigated the assessment literacy (AL) and practices before and after the pandemic. An online survey was administered to 300 educators from 57 countries to scrutinize their perceptions of the measures and correlations between their self-efficacy, AL and practices. The results revealed patterns of relatively controversial practices that could affect assessment quality. Weak to no correlations were found between AL and practices as the crisis itself may have mediated this relationship. Personal and contextual factors emerged in the qualitative data to reveal the limited effectiveness of some of these accommodations. Future research should lead to redefinitions of the AL concept to encompass flexibility to embrace adjustments to assessment frameworks and guidelines during crises.

Where’s my exam? Students’ perspectives of interrupted assessments.

M. Richardson[1]

[1]UCL Institute of Education, United Kingdom

Advances in practice might seek to promote student agency in assessment (Charteris and Smardon, 2019), but in reality (in England) students have little say in how they are assessed. Assessment remains something done to students rather than with students. Attempts to examine students’ understanding of assessment have been less about literacy and more about learning in general (e.g. DeLuca et al, 2018); or subject-specific knowledge (see Butler, Peng, and Lee, 2021). However, Smith et al. (2013) argue that assessment literacy requires understanding of the purpose and processes of assessment, and how these affect students’ perceptions.

We surveyed the “covid-cohort”: 17-18 year olds sitting A levels in 2022 following two years of disrupted schooling and little test taking preparation. Their experiences characterise test taking and being a test-taker: these unique insights provide some evidence about student assessment literacy in a time of challenge and change.

Through this study we seek to explore how students perceive the value of exams and make sense of their experiences of preparing for and taking them. We also hope to gather insight into the ways in which students gather information about exams, and how they make sense of this information both individually and as a community.

16:30 - 16:45

Did the pandemic expose a deficit in teacher assessment literacy in England – and is there a role for 'Powerful Assessment Knowledge'?

M. Johnson[1]

[1]Cambridge University Press & Assessment, United Kingdom

In this session I review data collected from 15 teachers in England over a 5-month period of the pandemic. During this time the teachers were responsible for assessing their students and submitting their grades for their end of school certificates, and this arrangement differed from teachers' practices in the past.

The teachers report wellbeing concerns, affected by heightened levels of assessment anxiety and workload. A combination of well documented weaknesses in assessment coverage in teacher initial training, a reduced role for universities in teacher preparation, and a lack of timely government guidance during this period of assessment change, suggest that the teachers were generally underprepared to deal with the assessment demands at this time.

The study teachers appeared to suffer from a transfer deficit, as their assessment literacy (AL) was inadequate for dealing with the changing professional conditions, which also included shifting social relationships. Consideration of the Powerful Knowledge concept allows insight into the characteristics of knowledge and expertise, and how it is acquired and incorporated into AL. It also allows discussion of the limitations of experience-based learning for professional teachers, highlighting the importance of ensuring that teacher development programmes enable them to encounter knowledge that has generalisable qualities.

Approaches to Assessing 21st Century Skills

16:00 - 16:15

Technology-Enhanced Assessment for and as Learning of 21st Century Skills in Maker Spaces: SkillTrack and Assessmake21

I. O'Keeffe[1]

[1]Learnovate, Trinity College Dublin, Ireland

This paper will present the findings from the ASSESSMAKE 21 Erasmus+ funded project which aims to provide, pilot, and validate novel assessment methods and tools for the assessment of 21st century skills. Pilot studies have been conducted in maker space settings in Ireland, Sweden, Greece, and Cyprus. As the maker movement is increasingly adopted into K-12 schools and non-formal makerspaces, students have more opportunities to generate unique, personalized artifacts, such as computer programs, robots or electronics, and develop 21st Century skills. However, assessment of these 21st skills is not easy, particularly within these open-ended environments where students create unique solution paths to problems, interact with peers, and act in both the physical and digital worlds. Furthermore, although there is a clear acknowledgement of the value of 21C skills, student understanding of these skills is often not addressed; there is a demand for assessment of the skills without the pre-requisite focus on the 'teaching,' 'learning' and 'doing' of the skills. The learning application SkillTrack was designed to address this challenge by supporting expectations around these skills and their development, cultivating skill literacy, cognition and reflection while creating an explicit, gamified, context for use.

Bridge21 - self reporting and reflection for 21st Century Skills

A. Bray[1], B. Tangney[1]

[1]Trinity College Dublin, Ireland

This paper will present the research conducted on analysing students' skills development when engaged in Bridge21 activities. Bridge21 is a pedagogical and activity model that focuses on learning by doing through collaborative teamwork, technology mediates and project-based learning activities. It has a strong focus on 21st Century skills development through its social constructivist approach to teaching and learning. These models have been used extensively both in the formal and informal education spaces where both the pedagogical and activity models have a strong emphasis on reflection, which is viewed as critical for student to take ownership and provide space for them to assess their own learning, as an individual but also as a team. As part of the out-of-school informal education programme that utilises the Bridge21 Approach, (i.e. the Bridge21 TY Programme), there has been extensive longitudinal research conducted on self-reporting assessment of confidence in skills development, which has been linked to students' well-being and educational attainment aspirations. Although the research is demonstrating positive outcomes, formalising and structuring an approach to assess these 21st Century skills remains challenging, particularly in the formal education system, however rubrics and student created artifacts are becoming common tools used for assessment in formal education.

CHARM-EU and Programmatic Assessment of 21st Century Skills

J. Byrne[1]

[1]Trinity College Dublin, Ireland

This paper explores the Programmatic Assessment approach used by CHARM-EU to assess student skills and competencies alongside traditional knowledge. CHARM-EU is a European University alliance formed of Trinity College Dublin, University of Barcelona, Utrecht University, the University of Montpellier and Eötvös Loránd University Budapest. As part of this alliance a novel Master's in Global Challenges for Sustainability was established. A central issue for "Global Challenges" is that they are inherently complex, ambiguous, and usually require diverse teams to work together to produce a solution. This means that transversal skills and competencies are crucial graduate attributes for this programme. Therefore, CHARM-EU utilises a Programmatic Assessment Approach to assess the students' knowledge and competencies under several Programme Learning Outcomes outlined in eight rubrics and evidenced by the students via an e-portfolio.
The assessment happens at multiple levels:

- Low-stakes assessment activities (i.e., Module assignments): focus on providing meaningful formative feedback such as a score in the rubric alongside written feedback.
- Intermediate stakes decision (i.e., Mentor advice): based on low-stakes assessment activities, serve to inform the student about its progress.
- High-stakes decision (i.e., Phase decision): pass/fail decision based on a holistic yet summative assessment of the low and intermediate stakes information.

Saturday,
12 Nov, 2022

Assessment Cultures IV

9:00 - 9:30

Back to learning outcomes. Changes in primary school assessment

G. Agrusti[1, 2], V. Damiani[1]

[1]LUMSA University, Italy
[2]Committee for Primary school assessment reform (Italian Ministry of Education), Italy

Recent changes in Italian regulations on how to express primary school students' assessment results led to a reconsideration of the role of learning outcomes. The study offers an overview of the main innovations introduced and the results of a set of trials carried out with a sample of over 800 primary school teachers on the correct formulation of learning outcomes. Among the main mistakes in selecting representative learning outcomes, there is the confusion between educational goals and observable instructional objectives, the use of verbs that require further clarifications either or the reference to too specific formulations, essentially reproducing the assessment task level.

9:30 - 10:00

Adapting to change: developing assessment cultures in English Primary Schools

E. Barrow[1], J. Golding[2], G. Grima[3], B. Redmond[4]

[1]Pearson Education, United Kingdom
[2]University College London Institute of Education, United Kingdom
[3]Pearson UK, United Kingdom
[4]Pearson, United Kingdom

Prior to the pandemic, English primary (aged 5-11) teachers worked with national summative assessments at ages 7 and 11, high stakes for schools if not for children. During the pandemic, such systems were replaced by locally-developed assessment and reporting. As part of our longitudinal (2018-22) study of primary teachers' and children's experiences of the current mathematics national curriculum, and their use of supporting resources, we evidenced changes in teachers' assessment practices over the course of the pandemic.

We had termly interactions with teachers in up to twenty schools, representative in a number of important characteristics. Data were derived from face-to-face and latterly, audio recorded, interviews, and teacher surveys, and analysed using an institutional ethnographic approach.

We show how many participants developed their mathematics formative assessment skills not only for distance learning contexts, but to deal productively with an enhanced range of prior learning brought back into classrooms as the pandemic abated. They often harnessed 'teacher-educative' resources to support their own learning about mathematics, mathematics pedagogy and their children's learning of mathematics, and applied that in pedagogically-challenging situations, and to productively inform successor teachers, reporting enhanced empowerment and professional trust. We discuss ensuing challenges as teachers return to external high-stakes assessments.

Educational Policy and Assessment IV

9:00 - 9:30

Assessing the impact of Applied General qualifications in England: Have outcomes for learners improved in the six years since the introduction of these qualifications?

H. Dalton[1], K. Mason[2], S. Nastuta[1]

[1]Pearson Education, United Kingdom
[2]Pearson, United Kingdom

This paper explores the extent to which Applied General Qualifications (AGQs) have improved outcomes for learners in England since their introduction in 2016. It draws on data and analysis from a large-scale survey of teachers and learners spanning the period 2018 to 2022, national Longitudinal Education Outcomes (LEO) data and descriptive data of the changing AGQ cohort.

AGQs were created as the result of a wider policy reform in England aimed at simplifying post-16 pathways and were specifically designed to equip students with transferable knowledge and skills needed to continue their education through applied learning. Six years on from the first teaching of these qualifications, this paper looks at research and analysis done to understand the impact that this reform has had on outcomes for learners.

The outputs from the survey are contextualised by emerging analysis using the LEO dataset which looks at outcomes for learners as they progress into the labour market and the extent to which there is evidence of impact from AGQs. Finally, the survey data is considered in the context of shifts in cohort characteristics over time and consider the extent to which this may affect outcomes and labour market impacts in the longer term.

GCSE and A-Level languages continuation in Northern Ireland: An analysis of the intersection of factors predicting young people's likelihood of opting-in to modern languages qualifications.

L. Henderson[1], J. Carruthers[1]

[1]Queen's University Belfast, United Kingdom

Against a backdrop of increased global importance of mastery in foreign languages, numerous major UK policy imperatives, such as the ‘Global Britain’ and ‘Levelling up’ agendas, fail to give adequate priority to language learning. Nonetheless, evidence consistently points to the potential benefits of linguistic competence for economic, social and cultural prosperity.

Our research examines young people's views and experiences of decision-making about participation in language study beyond the compulsory phase. The paper reports findings from a cross-sectional survey of young people in Northern Ireland (n=1278), to understand their motivations for opting-in to languages qualifications at two key transition points: from Key Stage 3 to GCSE study; and from GSCE to A-Level study. We use logistic regression modelling to examine the intersection of variables which predict positive continuations. In summary, we show that young people's decisions take account of a complex spectrum of demographic, structural, experiential and attitudinal factors: comprising socio-economic status, opportunities to learn, experiences and perceptions of formal assessment arrangements and attitudes towards the value of languages. We discuss the importance of our findings to policy formation, and argue that although their perspectives are a vital source of information in understanding attrition rates, they are not adequately considered.

10:00 - 10:30

What can a study of history tell us about our own assessment culture?

A. Watts[1]

[1]University of Cambridge, UK, United Kingdom

In this paper I will give a historic example in which a serious challenge was made to national examining in England and Wales. The challenge was that the marking of the markers was seen to be highly unreliable. Research by two scholars, Hartog and Rhodes, was published in the 1930s in publications which gained wide publicity.

Alongside the alarm that the research raised, a proposal was made that the judgement of examiners should be removed from examining altogether. What was needed was for exam questions to be ‘objective’, and so a greater use of a multiple choice format was proposed.

This kind of assessing was promoted in England by psychologist Philip Ballard whose book, The New Examiner, saw 14 reprints between 1923 and 1949. Objective marking was also recommended by academics from the United States who claimed that the examination boards in England were seriously out-of-date.

The paper will show how these proposals came into conflict with the assessment culture that already existed. It will explain how it was, rightly or wrongly, that the culture finally overcame the wholesale use of multiple choice questions. It will also ask how far this resulted from a relationship between teachers and examination boards.

National Tests & Examinations lll

9:00 - 9:30

The future of national tests – Comparing paper-based and digital assessments in upper secondary school mathematics

A. Lind Pantzare[1]

[1]Umea university, Sweden

In this presentation the results from a study where two separate test forms in mathematics, one paper-based and one digital, were administrated to two groups of students will be reported. The aim with the study was to learn more about differences in difficulty if mathematics items are served in a paper and pencil form or if they are administrated digitally where also the answers are given digitally. A second issue investigated was the cognitive workload and the use of scratch paper. The hypothesis was that when solving short answer items digitally students only use mental arithmetic and not the scratch paper and that a possible lack of scratch paper use might affect the proportion of students solving the item correct. A third issue is about writing down solutions to mathematical problems using an equation editor, which will be necessary when the national tests are digitalised in a couple of years.

New Visions for Assessment in Uncertain Times: Experiences of the Alternative Certification Model

S. Allan[1], S. Hill[1], E. MacIntosh[1], L. Wilson[1]

[1]SQA, United Kingdom

COVID-19 created unprecedented challenges for qualification systems worldwide. The Scottish approach for assessing National Qualifications in 2021 was termed the Alternative Certification Model (ACM), and involved a temporary move away from external assessment. The model used practitioner assessment and moderation, giving flexibility and autonomy to local education departments, schools and colleges. Support was provided by SQA through a national quality assurance exercise and subject-specific guidance and resources. The model was designed to deal with the challenges of COVID-19 disruption while ensuring learners were supported and awarded fair and credible qualifications.

SQA, Scotland's awarding body, carried out in-depth, mixed-method research to attempt to understand the experiences of learners and practitioners involved in the ACM. The results capture a complex picture, a range of experiences, and at times contradictory views.

Key findings include the challenges of reconciling practitioner flexibility with a shared national standard, competing aspects of fairness for individual learners and a consistent application of standards, and the increased workload for practitioners associated with such a model. Insights can help inform future thinking about education in Scotland, at a time of national debate on qualifications and assessment, and findings will contribute to ongoing considerations on how best to assess young people.

Fairness & Social Justice lll

9:00 - 9:30

The impact of the COVID-19 pandemic on the parity of academic and vocational 16-19 qualifications

J. Kaur[1], B. Ashworth[1], K. Mason[1]

[1]Pearson, United Kingdom

After completing their GCSEs, students in England have a choice of paths to take for their post-16 education. Some opt to take the academic, A-level, route, whilst others choose a vocational qualification, such as BTEC Nationals. The COVID-19 pandemic meant that exam series in 2020 and 2021 were cancelled and grades were awarded using teacher and centre assessed grades instead. This change led to increases in the proportion of students achieving higher grades in all qualifications. However, the increases in A level were larger than those seen in qualifications such as BTEC Nationals. This work looks at the changing relationship between attainment at GCSE and outcomes of selected A level and BTEC qualifications through the four summers since 2019. It suggests A level students with similar levels of prior attainment achieved higher outcomes when exams were cancelled than their BTEC peers and examines the extent to which this relationship is stabilising as we exit the pandemic. It discusses the implications for those BTEC students, who are more likely to be from disadvantaged backgrounds, and what can be done to restore parity on a level seen in 2019.

9:30 - 10:00

The comparability of grading standards in technical qualifications in England: how can we facilitate it in a post-pandemic world?

D. Rama[1], M. Walter[1], N. Zanini[1]

[1]Ofqual, United Kingdom

In England, Technical Awards are usually taken by 16-year-olds alongside academic qualifications and are available in a range of practical subjects (eg sport, health and beauty). Within the same subject area, several awarding organisations offer qualifications that are often used interchangeably. Yet, these qualifications are not designed to be equivalent to each other. Combined with the lack of a shared methodology used at awarding by different awarding organisations, this poses some risks to the comparability of grading standards in Technical Awards. These risks were heightened during the COVID-19 pandemic, when the external assessments that form a substantial component of these qualifications were replaced by teacher grades.

This presentation will report on the findings of quantitative research on the comparability of grading standards within the same subject across Technical Awards. We exploited a rich dataset on students' performance in these qualifications linked to a broad set of students' characteristics, including prior schooling and socio-economic background. We will use the findings to discuss the possibility of developing statistical indicators to be used at awarding to facilitate grading standard alignment, across qualifications and over time, as well as to reflect on how to ensure fairness to learners in a post-pandemic system.

Assessing and Evaluating the Impact of IB Career-related Programme Implementation in the County of Kent

T.N. Hopfenbeck[1, 2], S. Johnston[1], J. McGrane[1]

[1]University of Oxford, United Kingdom

[2]Norwegian University of Science and Technology, Trondheim, Norway

This paper presents the research study The Impact of IB Career-related Programme Implementation in the County of Kent, with the aim to evaluate and assess the implementation and the impact of the IB Career-related Programme (CP) in secondary schools in Kent. The research team designed a mix-method study to address the overall research questions What enhances implementation of the IBCP and what are the key challenges. Data was collected through interviews, document analysis, surveys and achievement scores. Interviews with staff demonstrated that capacity building and collaboration between schools were important factors for the success of the implementation, while student data showed that the programme offered skills which students found valuable for further studies and employment. Analyses of administrative data on student outcomes (n=379) from 18 schools offering the CP in Kent showed that most of the students (55%) went on to attend universities, 17% were in employment, 16% were engaged in apprenticeship, and less than 1% were unemployed after their CP. The findings of this study suggest that the CP provides the prospect of better life opportunities for a generation of students from less privileged backgrounds, some of whom have been the first in their families to attend university.

Psychometrics & Test Development IV

9:00 - 9:30

Creativity in examination question writing: how novel can examination questions really be?

F. Constantinou[1]

[1]Research Division, Cambridge University Press and Assessment, United Kingdom

The disruption caused by the COVID-19 pandemic has prompted calls for a new vision for assessment. However, any attempts at reimagining or reforming the current system should be informed by a comprehensive and in-depth understanding of foundational assessment processes. This study has focused on one such process, namely examination question writing, and has explored the creativity it entails. Creativity is important as it is a prerequisite for designing new questions, that is, questions that are different from those that have occurred in past papers. In contexts where the reuse of questions is not encouraged (e.g. England), creativity emerges as an important consideration. Drawing on the analysis of 3036 examination questions and interviews with seven professional question writers, this study sought to explore the boundaries of creativity in question writing. The study identified various factors that seem to restrict question writers' freedom, exposing a tension between creativity and constraint in question writing. This paper will present this tension and will argue that, despite the negative connotations that the notion of constraint typically carries, constraints are essential in question writing as they support the design of high-quality questions.

How Students Behave while Solving Critical Thinking Tasks in an Unconstrained Online Environment: Insights from Process Mining

A. Belyaeva[1], D. Federiakin[1], K. Tarasova[1], E. Orel[1]

[1]HSE, Russia

In the 21st century, it has become more popular to search data online rather than to use textbooks or other printed resources. However, it is not a trivial task for students to avoid false and not trustworthy information. This is why Critical Thinking (CT) is crucial on various levels of education. In the study, an instrument to measure CT was developed, using Evidence-Centred Design (ECD). The instrument is a test that consists of two parts, with a classical multiple choice and statement categorising items in the first one and a dilemma to be solved using an unconstrained online environment in the second. The purpose of this presentation is to analyse the process data from the second part of the assessment, analyse how students with different levels of CT behave on the Internet, and to come up with an explanatory framework for understanding students' critical thinking. To analyse the patterns, various process mining techniques (e.g. Alpha miner, Simple Heuristics miner, Inductive miner) were used in an open-source software ProM. The obtained results are going to be used for additional validity arguments for the framework.

10:00 - 10:30

Test taking strategies in multiple choice items – an analysis of a Swedish vocabulary test

C. Wikstrom[1], I. Laukaityte[2], M. Johansson[1], M. Nordvall[1]

[1]Umea University, Sweden
[2]Umeå university, Sweden

The Swedish Scholastic Aptitude Test (the SweSAT) is a standardized test used in the admissions to higher education in Sweden. This study focuses on group differences in performance in one of the SweSAT's subtests measuring word comprehension (WORD), to find out if male and female test takers differ in their strategies when selecting among the distractors. The analysis is based on SweSAT data from spring administration 2016 to spring 2019, and include 398 842 test takers, 140 WORD items and 700 response alternatives. Data is analysed using IRT and DIF analyses, where male and female test takers' response patterns are compared. Separate analyses are performed for the distractors to identify DDF, differential distractor functioning. The results show that there are differences between men and women that can explain some but not all the performance differences noticed. Male test takers tend to be more prone to take chances, but there are also differences in their choices of distractors, differences that vary with the test taker's underlying knowledge/ability. These results and potential explanations for deviating patterns and group differences in the choice of distractors are presented and discussed.

Assessment Cultures V

9:00 - 9:30

Exploring standards across assessments in different languages using comparative judgment.

L. Badham[1], A. Furlong[2]

[1]International Baccalaureate, United Kingdom
[2]International Baccalaureate, Netherlands

Awarding bodies face numerous challenges when trying to ensure fairness, validity and reliability in summative assessments. For global educational organizations such as the International Baccalaureate (IB), further challenges arise from the need to ensure comparable standards of assessments that are offered in multiple languages.

Traditional statistical techniques that compare scores across different language versions of tests can highlight potential comparability issues. However, such techniques are often intended for multiple-item assessments and are less suited for extended essay-based responses. The IB therefore conducted a study to explore whether comparative judgement could be used to compare standards across languages in literary essays. Bilingual examiners were asked to compare Literature essays in English and Spanish, before completing questionnaires to provide feedback on their experiences.

Findings from the study suggest that whilst high reliability could be achieved, examiner judgments were less consistent for bilingual pairings. Examiners' linguistic and professional experiences also had some influence on reliability. Feedback also suggested that English and Spanish Literature students approached literary analysis in fundamentally different ways, which raises questions about traditional definitions of inter-subject and intra-subject comparability. A new classification of "intra-disciplinary comparability" is proposed to accommodate assessment contexts where complex constructs manifest themselves differently in different languages.

Policy problems and paradoxes unveiled by the pandemic: Secondary grading and higher education admission in Norway

E. Hovdhaugen[1], S. Tveit[2]

[1]Nordic Institute for Studies in Innovation, Research and Education (NIFU), Norway
[2]University of Oslo, Norway

The Covid-19 pandemic has posed challenges to the policies and practices of secondary school grades and examinations, and the use of these assessment for admission to higher education in Norway. The cancellation of secondary school examinations in 2020, 2021, and 2022 prompted new discussions about the configuration of grades and examinations, with significant consequences for admission to higher education. This paper researches policy problems and paradoxes unveiled by the pandemic and how this disruption challenges inherited notions of fairness in the Norwegian assessment culture. The paper analyses two years of media discourses on secondary education grading and examination policies, as well as higher education admission policies from March 2020 through February 2022. The problems and paradoxes identified in this media discourse were then investigated in policy documents and research reports over the past ten years (2012-2022). The paper identifies unfairness in the competition between cohorts caused by grade inflation observable in Norway's national grading statistics for the 2020 and 2021 cohorts. Another issue called into question due to the pandemic which we discuss is the practice of assigning bonus points for choosing maths and science, which also inflates the GPA requirement for many academic programmes.

The Norwegian legacy of resisting formal grading: Paradoxes and dilemmas

S. Tveit[1], L.V. Sandvik[2], H. Fjørtoft[2]

[1]University of Oslo, Norway

[2]NTNU Norwegian University of Science and Technology, Norway

Grades' potentially harmful effects on learning have been increasingly emphasised in international research past decades and encouraged many educators to rethink the grading practice entirely. This has drawn attention to the assessment cultures in the Scandinavian countries, as this region already has a long-standing tradition of prohibiting formal grading in primary education. This paper explores the rationales that underpin the resistance to formal grading in the Norwegian assessment culture. The paper researches: (I) the arguments put forward in curriculum reforms for implementing and sustaining the policy of prohibiting formal grading in primary education, as well as for reducing formal grading in secondary education; and (II) the changes that can be observed in the assessment policy discourses related to formal grading in primary and secondary education from 1939 to 2020. The paper reports on a qualitative content analysis (QCA) of policy documents underpinning the curriculum reform of 1939, 1974, 1987, 1997, 2006, and 2020. The analyses illuminate teachers' roles in both supporting and certifying students' learning as a repeated tension throughout the 80 years of curriculum reforms. The paper discusses how the balancing of these tensions has taken different forms, as conditioned by the policy discourses at the time of reform.

Assessment against the backdrop of the Covid Pandemic V

9:00 - 9:30

Changing Assessment Culture

J. Behan[1], G. O'Sullivan[1]

[1]NCCA, Ireland

The International Education Assessment Network (IEAN) comprises international research and policy experts from similarly sized jurisdictions to share thinking from research and practice to help tackle long terms problems in educational assessment. The IEAN position paper Changing Assessment Culture is concerned with supporting systems to change assessment culture through reconciling beliefs, values and principles of high-quality teaching and learning, with ideas about assessment. The paper reflects on what is meant by ‘culture’ in broad terms before considering the concept of ‘assessment culture’. It reflects on some contextual factors relating to changing assessment culture and identifies processes conducive to supporting change. It presents an ecological view of assessment culture positioned within distinct contexts which requires an international, rather than global, perspective of education and assessment to be taken. The paper argues that realising change necessitates engagement with one’s sense of identity, including beliefs, values and emotions. It proposes that in complex education systems where change is ubiquitous, a culture of ‘learning’ must be a defining feature of professional practice. The paper contends that this culture of learning sets conditions whereby teachers are empowered to modify their practice and make professional decisions to support alignment between the purposes of education and assessment.

Pandemic predicaments and equity challenges: Emerging visions of assessment policy and practice for improving learning

A. Kanjee[1]

[1]Tshwane University of Technology, South Africa

This paper reviews the Ministry's response in mitigating the impact of the COVID-19 pandemic on schooling, and its implications on enacting new assessment policies and practices that foreground learning across fee and no-fee schools in South Africa. Data was obtained from policy makers, education role-players, teachers and a review of documents, and analysed using Cultural Historical Activity Theory. The findings revealed that while the pandemic had forced ministry officials into drastic responses, it also allowed for addressing long-standing policy challenges, and 'facilitated' new partnerships with key role-players. The findings also indicate that while the assessment revisions created conducive environments for teachers to transform their assessment practices, it also created additional tensions that impacted on their pedagogical practices. On the one hand, the reduced emphasis on summative assessments coupled with the promotion of formative assessments facilitated greater teacher agency in foregrounding learning over curriculum coverage and improving test scores. On the other hand, having less assessment information coupled with teachers' limited assessment capacity made teaching more challenging. Notwithstanding the capacity development implications, the extent to which these new practices and partnerships can instil a new vision of assessment among teachers, especially in no-fee schools, needs to be determined.

10:00 - 10:30

Pandemic and assessment: a powerful measurement tool

L. Heidmann[1], L. Neirac[1]

[1]DEPP - French Ministry of National Education, France

In France, a national assessment takes place every year since 2018 for all pupils at the beginning of their primary education: in September and January for first graders and in September for second graders. This paper-based assessment measures pupils’ academic performance in the basic skills of French and mathematics at three points of their schooling. In March 2020, due to the Covid pandemic, all schools in the country closed for at least three months, forcing students to continue learning alone, confined to their homes. We show that national assessment data is an extremely powerful tool to quantify the impact of this unprecedented crisis on student learning. From students' responses to a wide range of items, we construct composite indicators of their performance in the two disciplines using principal component analysis. We compare the learning progress of pupils who have experienced lockdown with the progress of pupils in first grade the previous year using difference-in-difference models. The results show that schools closure had a negative impact on pupils' progress in mathematics and especially in French. Moreover, this crisis has deepened pre-existing inequalities, as pupils entering primary schools with more difficulties were the most strongly affected by the closure.

Perspectives of End-users and the General Public on Assessment II

9:00 - 9:30

Sense and interpretability: Exploring educator's misconceptions when processing school performance feedback from large-scale assessments

E. Goffin[1, 2], R. Janssen[2], J. Vanhoof[1]

[1]University of Antwerp, Belgium
[2]KU Leuven, Belgium

In order for external (large-scale) assessment programs to thrive, end users need to be able to make sense of the output of such programs and perceive them as useful tools for school improvement. Feedback providers and test developers committed to optimizing feedback and its interpretability, need more insight into where their intentions diverge from actual user interpretations, where messages might get lost in translation, and where users' analyses and inferences might be invalid. Combining an information-processing perspective with a sensemaking/semiotic perspective, this paper investigates whether educational professionals grasp concepts central in school performance feedback, and interprets identified misconceptions in light of users' interactions with the data and strategies rooted in individual frames of reference. Teachers' and school leaders' engagement with authentic school performance feedback data from a national assessment is explored in a qualitative inquiry including a think-aloud procedure. Findings include that there is variability in users' conceptual understanding of feedback elements, that misconceptions arise in disconnects between explanatory paradigms, and that unfamiliar and complex concepts and presentations, particularly in terms of statistics, pose difficulties.

Irish Mathematics Teacher Perceptions of Classroom-based Assessments: Can Bridge21 Help?

A. Bray[1], S. Quigley[1]

[1]Trinity College Dublin, Ireland

This study explores the potential of a particular model of 21st Century teaching and learning – Bridge21 – to support Irish teachers' implementation of Classroom-based Assessments (CBAs) in the mathematics classroom. Within the context of the Irish curriculum, CBAs are a novel approach to assessment, and there has been considerable discourse in relation to their value and efficacy. The focus of this work however, is on teachers' perceptions of the Bridge21 model to support them in their implementation of CBAs in the mathematics classroom, exploring the positive effects of the structure provided by the model, the guided brainstorming session, and the focus on collaboration and peer learning. Results of this small exploratory study have shown that such an approach is of great benefit to the teachers, who had felt that the formal guidance and professional development provided by the state was somewhat lacking.

10:00 - 10:30

Trialling on-screen assessment: effects on student performance and experiences of students and teachers

J.M. Ryan[1], C. Balaban[1], Y. Bimpeh[1]

[1]AQA, United Kingdom

On-screen assessment is by no means a new concept, but the rise in online teaching and learning during the Covid-19 pandemic has accelerated the demand for reliable digital assessment. From April 2022 onwards, AQA will be conducting its first round of large-scale on-screen assessment pilots in a selection of schools throughout England. Digital exams will be delivered to Year 10 students (aged 14–15 years) in GCSE English Language, GCSE Mathematics and GCSE Sciences. Surveys and focus groups with students and teachers will be used to gather detailed feedback about how we can improve our on-screen assessments.

A move to digital assessment would represent a significant shift in the paradigm of high-stakes examinations, and at AQA we believe it is essential that students and teachers are actively involved in shaping this change. Initial feedback from members of AQA’s Student Advisory Group, who have trialled the pilot exams, revealed that overall they found the experience of taking the exams on screen to be positive and were excited about the possibility of digital assessment in the future.

This presentation discusses the findings from our analysis of students’ performance on these digital assessments and the feedback received from teachers and students about their experiences.

Keynote Speech

11:00 - 11:45

Assessment research: listening to students, looking at consequences

E. Smyth[1]

[1]Educational Research Centre, Ireland

This keynote address considers the kinds of information that should be used in looking at reform of assessment systems. The first part of the presentation focuses on the value of taking account of student voice in looking at the effects of different approaches to assessment, especially high-stakes examinations. While students provide invaluable insights, we also need to understand the consequences of different assessment approaches for educational inequality. The second part of the presentation discusses the extent to which different forms of assessment can reinforce (or indeed counter) social inequalities in student outcomes.

Keynote Speech

11:45 - 12:30

Content-Referenced Growth

D. Briggs[1]

[1]Research and Evaluation Methodology program, United States

In this presentation I will describe an approach to modeling the results from an educational assessment in a way that focuses attention on the qualitative distinctions in student learning that can be inferred from a quantitative measuring scale. This approach, which I call “content-referenced growth,” has four ingredients that require a significant investment in research and design: (1) a learning progression; (2) a cross grade scale; (3) item mapping; and (4) an interactive reporting system. The goal of content-referenced growth is to support interpretations of students’ scores relative to both the status of their understanding at one point in time, and their growth in understanding across points in time, relative to the content contained in the assessment. I introduce each ingredient and how they fit together in the context of newly developed learning progressions in mathematics and reading. I also discuss some preliminary results from piloting a prototype of an interactive reporting system with teachers who have experience administering and interpreting the results from the i-Ready Diagnostic, a large-scale assessment developed to support formative assessment purposes by the American company, Curriculum Associates.

Sponsors:

Fáilte Ireland An tÚdarás Náisiúnta Forbartha Turasóireachta National Tourism Development Authority

MEET IN IRELAND

AQA Realising potential

CiTO

GradeMaker

GA Partnership Moving tests online

TestReach

better examinations

Prodigy LEARNING

Adaptemy

State Examinations Commission Coimisiún na Scrúduithe Stáit

RM™

Questionmark powered by Learnosity

NCCA An Chomhairle Náisiúnta Curaclaim agus Measúnachta National Council for Curriculum and Assessment

Host Institutions:

Trinity College Dublin Coláiste na Tríonóide, Baile Átha Cliath The University of Dublin

Foras Taighde ar Oideachas Educational Research Centre

Main sponsors:

Pearson

CIVICA ideas into action

Coordinator:

easy CONFERENCES

ADVANTAGE event experts