

Assessing experimental science in 11–18 education: New research directions

Wednesday 12 October 2016 – Wellcome Trust, 09.30–19.00

Biographies of speakers



Professor Jens Dolin is Head of Department of Science Education at the University of Copenhagen. His research areas are teaching and learning science (with a focus on the development of competences) and organisational change (reform processes, curriculum development and teacher conceptions). He has participated in and been leader for a number of Danish and international science education research projects (including FP7-projects such as S-TEAM, Mind The Gap about inquiry-based science education and ASSIST-ME) and is member of a number of Danish and international boards and organisations.

After a long career as a high school teacher in physics and geography (involving development work and part-time teaching at university), he gradually drifted into research and employment in higher education and involvement in educational policy with emphasis on science education.



Professor Martin Fautley is Director of Research in the School of Education at Birmingham City University. After many years as a classroom music teacher, he then undertook full-time doctoral research working in the education and music faculties at Cambridge University, investigating teaching, learning, and assessment of classroom music making. His main area of research is assessment in classroom music, but he also investigates understandings of musical learning and progression. He regularly undertakes research and evaluation projects in the field of music education. He is the author of eight books, including *Assessment in music education*, published by Oxford University Press, and has written and published over fifty journal articles, book chapters, and academic research papers on a range of aspects of teaching and learning, and regularly presents the fruits of his researches at international conferences. He is co-editor of the *British Journal of Music Education*.



Emeritus Professor Patrick Griffin held the Chair of Education (Assessment) and directed the Assessment Research Centre at Melbourne University for more than 25 years. He was the Associate Dean of the Melbourne Graduate School of Education. He is one of six Australian Fellows of the International Academy of Education. He was a psychometric project team leader for UNESCO in southern Africa, and was awarded, in 2005, a UNESCO Research Medal by the Assembly of Ministers of Education from Southern African nations. He developed a system of teacher assessment signed into law by the Vietnam Government and applied to more than 380,000 teachers. He has led the development of leadership frameworks for the Australia Institute of Teaching and School Leadership. His work continues to focus on item response modelling applications to performance

assessment with a formative focus. He was the executive director of the Assessment and Teaching of 21st Century Skills (ATC21S) Project and is the lead editor of the ATC21S series. Volume 1 was published in 2012; Volume 2 in 2014 and Volume 3 is in progress for release in 2017. He is currently redeveloping that work with UNESCO to develop a global framework for a curriculum transition to future competences. In 2014 his work in linking assessment to teaching was published as *Assessment for teaching* by Cambridge University Press.



Dr Shaun Helman is The Transport Research Laboratory's Head of Transport Psychology, and is a cognitive psychologist who has been involved in researching road safety and driver behaviour for the last 15 years. He is particularly interested in the links between driver behaviour and safety outcomes such as collisions and injuries, and in high-risk groups such as young and novice drivers, those driving for work, and motorcyclists.



Professor Sir John Holman is an Emeritus Professor of Chemistry at the University of York, and senior education adviser to the Wellcome Trust and the Gatsby Foundation. His interests include educational policymaking, curriculum development, teacher development and practical science.

He is experienced as a secondary school teacher and head teacher and as a government adviser, and founded the National Science Learning Centre. He is a Trustee of the Natural History Museum and Chair of the Salters' Institute.

In 2014, John was awarded the Royal Society of Chemistry's Lord Lewis Award and the Royal Society's Kavli Education Medal. He was knighted in 2010 for services to education.



Dr Ian Jones is a lecturer in Mathematics Education at Loughborough University. Prior to this he was a Royal Society Shuttleworth Education Research Fellow, Secretary of the British Society for Research into Learning Mathematics, and has taught in various schools around the world. Ian is widely published in the discipline of mathematics education and sits on the editorial board of *Research in Mathematics Education*. His research interests include the assessment of mathematical understanding and the cognitive processes of learning to think mathematically. In particular, he leads a programme of research investigating the application of comparative judgement to the assessment of problem solving and conceptual understanding.



Dr Hilary Leever is Head of Education and Learning, leading work to improve science education – a named priority for Wellcome. This includes: an ambitious new campaign to improve science in primary schools, funding and advocacy for teacher professional development, and work bringing researchers and teachers together to build the evidence base to inform policy and practice.

After studying Natural Sciences at Cambridge, Hilary completed a DPhil at Oxford and progressed to an Assistant Professorial position at the Centre for Molecular and Behavioural Neuroscience at Rutgers, USA. Her research covered children's reasoning and learning strategies, early cognitive and language development, and developmental disabilities. After returning to the UK, she joined the Campaign for Science & Engineering as Assistant Director and worked on a breadth of policy issues, including leading on education, before joining Wellcome in 2011. She is a governor of a 5–16 comprehensive school.



Professor Thomas McLeish FRS is a theoretical physicist whose work is renowned for increasing our understanding of the properties of soft matter. This is matter that can be easily changed by stress, including liquids, foams and biological materials.

Although Tom's work is mostly theoretical, he also works closely with those performing experiments and in industry. He has made significant advances in modelling the structure and properties of complex entangled molecules, blends of substances that don't usually mix (like oil and water), and modelling micro-structures under flow. This allows us to predict more easily complex fluid behaviour and processing in an industrial setting.

Tom is Director of the Durham Centre for Soft Matter at Durham University, a multidisciplinary team that works across physics, chemistry, mathematics and engineering. Tom's other interests include historical studies of medieval science, and he is a member of the Institute of Medieval and Early Modern Studies at Durham and author of *Faith and wisdom in science* (2014).



Dr Eilish McLoughlin is the Director of the Centre for the Advancement of STEM Teaching and Learning (CASTeL) and a Senior Lecturer in the School of Physical Sciences at Dublin City University. Her research interests are focused on STEM teacher education, at both pre-service and in-service levels, as well as use of innovative pedagogy and assessment at second and third level. She played a leading role in two major pan-European projects focused on teacher education in Inquiry Based Science Education (IBSE) – as coordinator of the FP7 ESTABLISH project (2010–2014) and as a member of the coordinating team of the FP7 SAILS project (2012–2015). She is a member of several international boards in science education, including ESERA, IUPAP Physics Education, Science on Stage and is current co-chair of the Institute of Physics (IOP) in Ireland Education group.



Dr Ronny Scherer currently works as a postdoctoral researcher in the Centre for Educational Measurement at the University of Oslo (CEMO) in Norway. His research mainly focuses on the development and evaluation of computer-based assessments in the domains of problem solving and science. He works on ways to measure complex constructs that play a critical role in education (eg adaptability, critical thinking, and perseverance). In order to study which factors determine the development of such constructs, he also focuses on teachers' instructional practices, attitudes, self-beliefs, and the integration of modern technologies in classrooms. Ronny obtained his PhD in science education from Humboldt-Universität zu Berlin in Germany.



Joanne Sim is the School Assessment Design & Development Coordinator for the New South Wales (NSW) Department of Education, which currently delivers online diagnostic science tests to over 100,000 students in Years 6, 8 and 10. In 2014 and 2015, she led the Australian writing team for the Collaborative Assessment Alliance research project into assessing collaborative problem solving in an online environment.

Joanne is an experienced and innovative science teacher with extensive involvement in assessments, curriculum development and resource production for the NSW Department of Education, NSW Board of Studies and Science Teachers Association of NSW.



Professor Kay Stables is Professor of Design Education at Goldsmiths, University of London. A founder member of the Technology Education Research Unit (TERU), she has directed and contributed to projects in primary and secondary education in the UK and overseas. With Richard Kimbell, she authored the TERU retrospective; *Research Design Learning* (2007). More recently, together with Steve Keirl, she published *Environment, ethics and cultures: Design and Technology education's contribution to sustainable global futures* (2015), and *Critique in Design and Technology education* (in the press) with P. John Williams from Curtin University, Australia. Her recent research has focused on design, creativity and sustainable development, digital tools in assessment (the e-scape project) and designerly well-being. Current research includes creating dialogic frameworks for supporting the development of design and technology capability in digital environments.



Jennifer Stafford-Brown works as a consultant for a leading exam board in qualification development and external assessment design. The main focus of her work is within the fields of sport and exercise sciences and physical education. In addition Jennifer holds the positions of Chief Examiner across level 2 and level 3 vocational qualifications for external assessments and Senior Standards Verifier for internal assessments. She also works as a Subject Expert for Ofqual in qualifications accreditation.

Jennifer has authored over 20 textbooks as well as digital resources for organisations such as BBC Bitesize and Channel 4 learning. She has over 20 years' experience of working as a practitioner and educational consultant within the secondary school and further education sector.



Dr Chris Wheadon is the founder of nomoremarking.com. He has worked in the past as a psychometrician for the CEM centre at Durham University and was Head of the Research Unit at AQA.



Professor Mark Wilson is a professor of Education at UC, Berkeley, and also at the University of Melbourne. He received his PhD degree from the University of Chicago in 1984. His interests focus on measurement and applied statistics, and he has published over 100 refereed articles in those areas, and over 50 invited chapters. Recently he was elected President of the Psychometric Society, and is currently President of the US National Council for Measurement in Education (NCME); he is also a Member of the US National Academy of Education, a Fellow of the American Educational Research Association, and a National Associate of the US National Research Council. He is Director of the Berkeley Evaluation and Assessment Research (BEAR) Centre. His research interests focus on the development and application of sound approaches for measurement in education and the social sciences, the development of statistical models suitable for measurement contexts, the creation of instruments to measure new constructs, and scholarship on the philosophy of measurement.

Assessing experimental science in 11–18 education: New research directions

Wednesday 12 October 2016 – Wellcome Trust, 09.30–19.00

Abstracts of speakers' presentations

Professor Jens Dolin

'Improving the assessment of experimental science through evidence-based research: key questions'

The presentation will look at summative assessment of experimental science in the context of science education. It will point at the necessary alignment between the purposes of experimental science in school, the way experimental science is realised, and the assessment forms used for summative (as well as formative) purposes. Based on research findings it will give some directions for future practice that can combine more valid summative assessment with improved student learning.

Professor Martin Fautley

'Assessment in music education: some current issues'

In this presentation, Professor Fautley will outline and address some of the issues of concern to music teachers with regards to assessment in the classroom. He will discuss the ways in which the demands of school-based assessment tracking systems can sometimes lead to unintended consequences in the music classroom. The presentation will also look into the issue of 'unilinear' progression, and how this can be a problem for music education, as well as how tracking systems can sometimes impede logical assessment protocols in the music classroom.

Professor Patrick Griffin

'Assessment in a knowledge economy: new approaches and learning analytics'

This presentation examines pressures building on education as the world enters a global knowledge economy. Not for the first time are education systems under pressure to change: the Internet is becoming the source of knowledge and assessments; the World Wide Web makes assessment information available in real time; complex interactive tasks involving multiple contributors are becoming available. Education has to deal with the impact of changed expectations at all levels of education. The 4Rs – Reading, wRiting, aRithmetic and Remembering – of an industrial curriculum respond to the 4Cs of a competency curriculum – Creativity, Collaboration, Critical thinking and Communication.

Assessing the new competences requires a rethink of educational measurement. The contributions of individuals to teams and of teams to workplace outputs require new forms of assessment and reporting: digital technology and data analytics offer new ways to interpret performances; teacher education and employment are based on older premises; government wants to support change but fears profound change; parents can't recognise the scores or grades in today's assessments and reports and worry

about their children's future; workplace outputs are changing from products to information, but vacancies are not being filled with work-ready employees. Education is, on the one hand, the cause of consternation while, on the other hand, is regarded as the salvation to prepare people for work, life and learning. Can education succeed with revised, digital age opportunities in assessment and reporting?

Dr Shaun Helman

'Measuring the impact of policy changes for young and novice driver safety in GB'

Between the late 1980s and early 2000s (and continuing to the present day in more sporadic forms) a programme of work at The Transport Research Laboratory examined the safety (or lack of it) of young and newly qualified drivers. The work began by trying to understand why such drivers are at such an inflated risk of crashing, when compared with older and more experienced drivers. It then progressed to examining various interventions that might be applied (for example, through the licensing process) to increase safety. This short talk will outline the major landmarks in this research programme, including the creation of the UK's Hazard Perception test (which drivers need to complete in order to gain their licence) and the subsequent assessment of its impact on self-reported collisions. Current research directions and projects will also be described.

Dr Ian Jones

'Comparative judgement for robust assessment: the case of mathematics'

Mathematical knowledge and facts are easy readily assessed using traditional test questions. Deep knowledge of mathematics, commonly called conceptual understanding, is harder to define and to assess. Recently a novel application of comparative judgement to the case of mathematics has overcome this problem. Students are presented with a short prompt, such as 'What is a derivative?', and write a one-page open-ended response. Assessors then undertake pairwise decisions as to which student is 'the better mathematician'. The decision data are then modelled to generate a score for each student. A programme of research led by Loughborough University has shown this approach yields valid and reliable measures of students' conceptual understanding across a variety of mathematical concepts, ages and contexts.

Dr Eilish McLoughlin

'Collecting evidence of inquiry learning in the science classroom'

This contribution will discuss how we can support teachers in changing their assessment practices so as to recognise and value the skills and competences that can be developed through inquiry learning in science at second level. Several key questions need to be considered by teachers in planning their assessment practices, in terms of what to assess, how to collect evidence, when can this evidence be collected and who assesses? In particular, teachers should consider the purpose of formative assessment and how they will use evidence of student achievement to make decisions about next steps in instruction as well as to provide feedback to students on their understanding of scientific knowledge and development of skills and competences.

Literature and outcomes of several projects have highlighted the impact of professional development programmes that supported teachers in adopting inquiry-based approaches in their classrooms, eg

organising experimental work so that students raise questions, develop their own hypotheses, test their ideas critically and suggest further questions. However, further research is needed to determine how professional development programmes can be designed to support teachers in adopting formative assessment practices and the influence of their beliefs about student learning and their assessment literacy on these practices.

Dr Ronny Scherer

‘Science assessments of the 21st century: current innovations and challenges in computer-based testing’

In our knowledge- and information-rich society, students need to develop transferable competences that can be applied in both domain-general and domain-specific contexts. Among other skills, the abilities to think critically, retrieve and evaluate information from multiple sources, collaborate, solve ill-structured problems, persevere, and adapt to novel, changing, and uncertain situations are critical to deal with the complex, real-life issues and challenges. It has therefore become the designated aim of education to help students to develop these skills. At the same time, this aim has called for assessments that capture the nature of these complex skills.

This presentation addresses the potential and the challenges of developing such assessments with a special emphasis on the role of computer-based testing for science education. Assessment and modelling approaches are exemplified based on the concept of complex problem solving. Finally, the potential of process data that can be obtained from computer-based assessments is emphasised, and future directions of the measurement of 21st century skills are proposed.

Joanne Sim

‘Online Practical Component for ‘Validation of Assessment for Learning and Individual Development’: a New South Wales Department of Education innovation’

The New South Wales (NSW) Department of Education has developed and implemented the Online Practical Component (OPC) to assess the practical outcomes of the science syllabus through an online multimedia environment. The OPC is now incorporated into the full cohort online multimedia interactive diagnostic science assessment known as Validation of Assessment for Learning and Individual Development (VALID). The presentation will include a live online demonstration of the OPC and other test items created in the online environment to assess the practical skills for science. Lessons learned by NSW during the transition to online assessments will also be outlined.

Professor Kay Stables

‘Capturing capability: digital approaches to authentic performance assessment in Design and Technology’

This presentation outlines our approach to authentic performance assessment, starting with research for the Assessment of Performance Unit (DES) for whom we developed test activities that enabled learners to produce portfolios in response to tasks in Design and Technology (1985–91). We extended this approach to paper-based summative assessment portfolios in a QCA project to assess design innovation at GCSE (2002–04). Through further development we created digital portfolios in the e-scape project

(2004–09). This enabled learners to create real-time, multi-media Web-portfolios in response to design, science and geography tasks, in controlled assessment environments. This presentation will explain how and why we developed Adaptive Comparative Judgement (ACJ) as the principal tool for assessment and how research has shown its reliability for summative assessment and its formative value for teachers and learners. Finally, I will touch on current research creating on-screen avatars that support learners' reflection when designing, thereby creating further evidence of their thinking.

Jennifer Stafford-Brown

'Practical assessment in GCSE Physical Education and vocational sport qualifications'

In GCSE Physical Education, 60% of the final grade is based on practical assessment. However, the grades achieved in practical assessment consistently show bunching at the top end with very narrow ranges between grade boundaries, which has clear implications for measurement accuracy.

As regulatory changes have required increased scrutiny in vocational practical assessment, steps have been taken in vocational sport qualifications to meet this challenge, for example through the use of audio-visual footage as part of the practical assessment process. Issues around the introduction of this technology, such as budget and staffing implications, will be discussed.

Lessons learnt from these qualifications can be applied to other sectors to improve the quality and value of summative practical assessment for student achievement.

Dr Chris Wheadon

'Sharing standards: how comparative judgement can help'

Chris Wheadon from nomoremarking.com will be talking about his experiences in using Comparative Judgement (CJ) for assessment.

In a CJ exercise, instead of applying a mark scheme to evaluate a candidate's work, a judge compares pairs of scripts (or other 'work products') and simply has to decide which is better. By aggregating lots of these paired comparison judgements across judges and scripts, it is possible to create a single rank order for all the scripts. Chris has developed software that allows these judgements to be made and analysed online. He has worked with a range of assessment organisations including Ofqual, Cambridge English, Ark Academy Chain, Eton College, the Fischer Family Trust, and the ARD Research Division. Comparative Judgement at NoMoreMarking.com has been used in scenarios as diverse as measuring pupil progress in schools, peer assessment, assessing creativity, measuring item difficulty and monitoring standards over time.

Professor Mark Wilson

'Roadmaps to help develop assessments for learning progressions in science'

This talk will discuss two examples of how we at the BEAR Center in Berkeley are creating 'roadmaps' of student progress in science using a variety of assessment approaches. One example is in an area that is a traditional science content—the Atomic Molecular Model—and one that is relatively new for assessment, a practice – Argumentation. The two examples are also used to illustrate how the BEAR

Assessment System can help develop these roadmaps of student thinking, and lead to assessments that focus on specific diagnostic steps along the way. The talk will conclude with some thoughts about the challenges and opportunities that this approach can generate.