

# Oxford University Centre for Educational Assessment



## Annual Report 2012-13

June 2013

Jo-Anne Baird, Therese Hopfenbeck, Daniel Caro and Joanne Hazell



## Table of Contents

Foreword.....	3
Director's Overview .....	4
The Centre.....	6
Research Themes .....	6
Governance .....	7
Staff.....	8
Research Students.....	8
Departmental Associates.....	10
Honorary Research Fellows .....	10
Research Visits.....	11
Research Programme.....	12
Projects.....	12
Publications in 2012-13.....	15
Research dissemination.....	17
OUCEA Events .....	18
Student Assessment Network .....	21
Committee Membership and Expert Groups .....	22
Media.....	22
Future Plans.....	23
Students.....	23
Research Programme .....	23
Appendix: Abstracts .....	24

## Foreword



I am not involved in the day to day work of the Centre, but I get a good insight into its progress from my position as Chair of the Joint Advisory Board. From my various Oxford perspectives, principally as Warden of Keble College and currently as Chair of the Conference of Colleges, and my previous experience in government, I recognise the importance of the research being undertaken and reported on in the following pages. The Centre offers an independent space in which researchers can develop new ways of thinking about the fundamentals of assessment.

In the period since the last report, there have been many achievements. In particular, the Centre's capacity has increased significantly. There are now three academic members of staff, supported by an administrator, and eight doctoral students. And the range and number of projects being undertaken has also increased with support from various funding bodies, including governments, inter-governmental organisations and the private sector. The projects reflect the Centre's overall strategy of focusing its research on both the technical features of assessment and the inter-action between assessment and society in the broadest sense. As the report makes clear, the outcomes of this research are being widely communicated in specialist journals and in seminars and workshops held in Oxford and elsewhere. In other words, impact is taken very seriously.

I am clear that the Director and her team are in no sense complacent about what has been achieved so far. The Centre is still in the early stages of an exciting journey with many challenges ahead. But there is much in this report of which to be proud.

A handwritten signature in black ink, which appears to read 'Jonathan Phillips'.

Sir Jonathan Phillips

Warden of Keble College, Oxford University

## Director's Overview

The Oxford University Centre for Educational Research Centre was pleased to host a visit from Dame Marjorie Scardino, Chief Executive of Pearson, in October 2012. Dame Marjorie reiterated Pearson's commitment to independent research and was positive about developments since the Centre was launched by her in 2008. Dame Marjorie retired in December 2012 and John Fallon has taken on the post of Chief Executive of Pearson.

OUCEA has grown from 25 to 30 people. We made an excellent appointment in Daniel Caro to a new Research Fellow position. Daniel brings a wealth of expertise in quantitative methods and has been publishing in high quality journals on issues such as the impact of socioeconomic status upon test scores. The number of research students has also increased, with new doctoral projects having begun on the relationship between learner identity and assessment, the use of test data by school Principals in Indonesia and the impact of taking examinations at an early age. I am also pleased to announce two new Honorary Research Fellows: Dr Michelle Meadows (Director of the Centre for Educational Research and Policy, England) and Dr Edward Wolfe (Pearson Assessment & Information, US).



Professor Jim Tognolini, Dame Marjorie Scardino and Professor Jo-Anne Baird at OUCEA

Our research dissemination is intended to influence a broad audience of academics, practitioners and policy-makers. As such, our research has been cited by the press and in government documents. We held policy-focused events, as well as seminars, CPD activities and workshops and spoke at 15 external events. Two of our symposia on examination reform were held in conjunction with other bodies – the Advisory Committee on Mathematics Education and the Oxford Education Society. A number of publications have come to fruition arising from previous events hosted at the Centre and from past projects this year. Four Special Issues of academic publications were published relating to our research programme. It was a privilege this year to give a St. Anne's College Domus Seminar and an opening address for the Centre for Educational Measurement at the University of Oslo.

Over the past year, we have been engaged with a number of research projects funded by the OECD, the Crown Prince of Bhutan, Pearson and the State Examinations Commission for Ireland. In each case, the issue being researched is one for many societies internationally and as researchers we have the challenge of ensuring that our work builds upon previous work whilst ensuring that we know enough of the context to understand how findings might differ in a new setting. These projects, on policy, quality of marking and the predictability of examinations, deal with fundamental aspects of how assessment articulates with societies and their values.

In January 2013, I took on the role of President of the Association for Educational Assessment-Europe. It is an honour to contribute to the Association in this way and I have been working with colleagues across Europe, particularly on the Association's Council, to grow the organisation's membership and scientific programme. The [14th Annual Conference](#) is being held at the Sorbonne in Paris on 8 – 10 November 2013 and the theme is *International surveys, policy borrowing and national assessment*. Our keynote speakers are Nathalie Mons (Université de Cergy-Pontoise), Paul Newton (Institute of Education, University of London), Andreas Schleicher (OECD) and Pierre Vrignaud (Université de Paris 10).

*Jo-Anne David*

Pearson Professor of Educational Assessment

Fellow of St. Anne's College



# The Centre

Our Centre is composed of academic staff (3) and an administrator, who are directly employed by OUCEA. Additionally, there are research students (8), Departmental Associates (7) and Honorary Research Fellows (11). This makes for a vibrant research environment where people encourage academic excellence and risk-taking.



## Research Themes

The research themes, which guide the Centre's work, were revised and agreed by the Centre's Joint Advisory Board in June 2012. The Centre's overall aim is to investigate the impact of educational assessment. Connections between policy, assessment theory and practice need to be elucidated, researched and critiqued for assessment to play a better role in improving education.

The Centre's main research priorities are to:

- Conduct research into the role of assessment in developing, implementing and monitoring **system-level change**.
- Develop **models of educational assessment** to increase understanding of learning processes, resource use and outcomes in education systems.
- Undertake systematic analyses of **assessment, public policy and the interaction between them** in various sectors of education and subject domains. For example, the use of international test data within and across education systems.
- Investigate **how assessment is perceived and used by stakeholders** (learners, parents, institutions, employers, policy-makers) within education systems.
- Critically examine notions of **validity, reliability and standards** in relation to different modes of assessment and to different purposes.
- Evaluate the potential for **the use and application of psychometrics** to enhance the validity and reliability of educational assessment.
- Explore the opportunities for, and the impacts of, **new assessment technologies** in test design, test distribution, test administration and results delivery.



## Governance

The Centre holds Management Group meetings termly, with the following representatives:

- Department of Education Director
- OUCEA Director
- Department of Education Administrator
- OUCEA Administrator

In addition, we hold an Advisory Board meeting twice a year, with the following members:

- Sir Jonathan Phillips, Warden, Keble College (Chair)
- Mark Anderson, President, Strategy and Business Development, Pearson International
- Professor Jo-Anne Baird, Director, Oxford University Centre for Educational Assessment
- Professor Richard Daugherty, Honorary Professor, Cardiff University and Senior Research Fellow, Oxford University
- Professor John Gardner, Deputy Principal (Education and Students), University of Stirling and Visiting Professor, Oxford University
- Professor Anne Edwards, Director, Department of Education<sup>1</sup>
- Professor Ernesto Macaro, Director, Department of Education
- Dr Kathleen Quinlan, Head of Educational Development, Oxford Learning Institute
- Professor Jim Tognolini, Senior Vice President, Pearson Research and Assessment, Pearson International and Senior Research Fellow, Oxford University

Kathleen Tattersall, who was a hugely supportive member of our Advisory Board sadly died of cancer on 23 January 2013. We miss her engagement with our research and her sage advice on our strategic direction.



---

<sup>1</sup> Anne Edwards was a member of the Advisory Board in her capacity as Director of the Department of Education. Ernesto Macaro took over that role in January 2013.

## Staff



**Jo-Anne Baird's** research interests include examination standards, policy and systemic aspects of assessment, e-assessment and human judgment in assessment. Her background is in academia (Bristol; Institute of Education) and in industry (Head of Research at the Assessment and Qualifications Alliance). Jo-Anne is a Visiting Professor at Queen's University, Belfast. College: St. Anne's



**Therese Hopfenbeck** conducts research on international tests, especially the relationship between student motivation and results on the tests. She is also interested in classroom assessment and self-regulation. Therese has recently been appointed by the Norwegian Minister of Education to chair the board of the Norwegian Centre for Learning and Behavioral Research in Education. College: St. Anne's



**Daniel Caro's** primary research interests reside in the areas of educational inequality, international large-scale assessment (LSA) studies, mixed models in cross-sectional and longitudinal settings, and causal inference with observational data. He is also developing a package in R ('intsvy') for analysing international LSA datasets. Package 'intsvy' allows users to import, merge, and perform statistical analysis with data from PISA, TIMSS, and PIRLS while taking into account plausible values and replicate weights.



**Joanne Hazell** is the OUCEA Administrator and supports all of the activities of the Centre. Jo deals with contractual and budgetary matters, as well as research proposal processes, personnel matters, organisation of events and support for students and visitors. She also maintains the website. Jo works part-time at the Centre and is assisted by doctoral students with some of these activities.

## Research Students



**Victoria Elliott** graduated with a DPhil from the Centre in 2011. Her supervisor was Professor Ingrid Lunt and her research concerned **decision-making and judgment in assessment**. Victoria went on to a post teaching at Warwick University and this year has taken up a post as a Lecturer in Education at York University.

There are currently eight research students at the Centre.



**Carol Brown** is researching the relationship between **expectations, values and A-level achievement** using a mixed methods approach. This research draws upon her professional background in social work and a 15 year career as a psychology teacher. She has also authored a number of psychology A-level and undergraduate textbooks. This year she has presented her work at the Student Assessment Network seminars and the Department's Student Conference (STORIES) and is an active participant of the Wolfson College Mind, Brain and Behaviour Research cluster. Supervisors: Dr Therese Hopfenbeck and Professor Jo-Anne Baird. College: Wolfson



**María Teresa Flórez** is conducting research on the topic of **assessment policy reform in Chile**. She is conducting an historical and contemporary analysis of policy-making systems and actors using polysystems theory. Previously, María Teresa worked in policy implementation at the national level, with a focus upon assessment for learning. In 2012 María Teresa won the prestigious Association for Educational Assessment-Europe's New Researcher Award and gave a keynote presentation at the annual conference of the



association in Berlin on 8-10 November 2012. She also presented at the ECER main conference in Cadiz, in the Education Policy Network. Supervisor: Professor Jenny Ozga. College: St Cross.



**Malcolm Hayes** is researching vertical **equating in national curriculum tests** in England. Malcolm has been professionally involved in standard-setting for the national curriculum tests in mathematics for over a decade and has worked with colleagues from the Centre this year on the Pearson-funded project researching marker reliability. College: St Anne's.



**Yanhong Li** is researching **grade inflation in higher education in the UK**. She has worked at Oxford Brookes University, Middlesex University and Portsmouth University and published on issues around performance measurement and assessment in both the public and business sectors. Supervisor: Professor Ken Mayhew. College: St Anne's.



**Manzil Maqsood** has constructed software to support teachers' use of **assessment for learning in classrooms in Pakistan**. The software is developed for Year 5 mathematics curriculum followed in Pakistan. She intends to investigate if teachers and students in primary schools of Pakistan will accept and use this software in their daily schooling routine. She is currently conducting fieldwork. Manzil's professional background was in teaching computer engineering and software project management in Lahore, Pakistan. Supervisors: Professor Jo-Anne Baird and Dr Gabriel Stylianides. College: St Antony's.



**Yasmine El Masri** started her DPhil in October 2010. She is particularly interested in **comparability issues in international assessment**. Her current research compares various language versions (English, French and Arabic) of the same PISA and TIMSS science tests in terms of their difficulty and demands. She is comparing data of three different countries: England, France and Jordan. Yasmine is investigating the usefulness and limitations of psychometric techniques, as well as value judgement methods in analysing such data. Her tentative title is 'Does the Language of Science Matter?' Yasmine completed the online course offered by the University of Western Australia – Introduction to Rasch Measurement of Modern Test Theory – with a high distinction. Yasmine is Student coordinator of AEA-Europe. She organised the pre-conference doctoral workshop in Berlin, November 2012, and is involved in the organisation of the annual conference in Paris 2013. Yasmine has also been Assistant Editor of *Assessment in Education: Principles, Policy and Practice* since November 2012. Supervisors: Professor Jo-Anne Baird and Dr Jane McNicholl. College: St Anne's.



**Agni Paramita** is researching **schools' use of National Exams data in Indonesia**. Using case studies and principal and teacher surveys, the study seeks to explore the role of leadership, district support, and professional opinion of the assessment data in this process. Agni is the Student Assessment Network Coordinator for OUCEA. Supervisors: Professor Jo-Anne Baird and Professor Pam Sammons. College: St Anne's.



**Rachel Taylor's** research is exploring the issues surrounding **early entry to GCSE mathematics**. Specifically, she is interested in the effects of early entry on teaching practices within schools and the consequences of this, particularly in relation to student performance. Rachel is studying part-time and is also employed as a researcher by the Centre for Education Research and Policy, affiliated with the Assessment and Qualification Alliance awarding body. Supervisors: Professor Steve Strand and Professor Jo-Anne Baird. College: Kellogg.

## Departmental Associates

Members of the Centre from the Department of Education and their research areas are:

- Dr Chris Davies – ICT in education
- Dr Jane McNicholl – science education
- Professor Jenny Ozga – policy in education
- Professor Pam Sammons – accountability and management
- Professor Judy Sebba – policy, assessment, learning
- Professor Steve Strand – equity in assessment
- Professor Kathy Sylva – early years and children's development

## Honorary Research Fellows

Visiting Professors, Senior Research Fellows and Visiting Associates of the Centre are:

- Professor David Andrich (University of Western Australia) – Rasch measurement
- Professor Richard Daugherty (formerly University of Aberystwyth and currently Honorary Professor, Cardiff University) – assessment for learning and assessment policy
- Professor Jannette Elwood (Queen's University, Belfast) – gender and children's rights in assessment
- Professor John Gardner (University of Stirling) – assessment for learning and teachers' understanding of assessment
- Professor Art Graesser (University of Memphis) – cognitive science and large-scale assessment (he is currently serving as Chair of the Framework group in PISA Collaborative Problem Solving 2015)
- Dr Michelle Meadows (Assessment and Qualifications Alliance) – assessment reliability and validity, qualification design, education policy
- Professor Gordon Stanley (University of Sydney) – public examinations
- Professor Gordon Stobart (Institute of Education, University of London) – assessment for learning
- Professor Jim Tognolini (Pearson International) – Rasch measurement, standard-setting and public examinations
- Dr Ed Wolfe (Pearson Assessment & Information) – measurement, rater effects

Lucy Simpson is a research associate at the Centre. Lucy is a doctoral student at the University of Bristol researching trust in assessment.



## Research Visits

OUCEA hosted research visits from the following academics:

- Associate Professor Christian Brandmo, University of Oslo, Norway
- Professor John Gardner, University of Stirling, Scotland
- Dr Monika Holmeier, University of Applied Sciences and Arts Northwestern Switzerland
- Associate Professor Marit Kjærnsli, University of Oslo, Norway
- Dr Joshua McGrane, University of Western Australia

# Research Programme

## Projects

Research projects underway in the Centre in the period of this report are detailed below.

### *Assessment policy processes in Norway*

Funder: Organisation for Economic Co-operation and Development; Principal Investigator: Dr Therese Hopfenbeck (OUCEA); Collaborators: María Teresa Flórez, Yasmine El Masri (OUCEA) and Assistant Professor Astrid Tolo (University of Bergen).



Professor Sølvi Lillejord, Associate Professor Astrid Tolo, Dr Therese Hopfenbeck and Professor Jenny Ozga at a Department of Education public seminar, 2013



Dr Therese Hopfenbeck presenting at AEA-Europe conference, Berlin 2012

The OECD-funded project *Governing Complex Education Systems* explores which different governance mechanisms and knowledge options facilitate effective steering of complex education systems.

Therese was commissioned by the Norwegian government and OECD to conduct the Norwegian case study in this project. Similar case studies have so far been conducted in the Netherlands and Poland.

This Norwegian case study investigated how the educational programme *Assessment for Learning* – a four-year programme (2010-2014) involving more than 400 schools – has been implemented in Norway.

Preliminary analyses were presented for the Ministry of Education in Norway in March 2013. The report has been submitted to the OECD and it is due to be published on the OECD website in June 2013.

## ***World Class Qualifications Review: Research evidence relating to proposals for reform of the GCSE***

Funder: Pearson UK; Principal Investigator: Professor Jo-Anne Baird; Collaborators: Dr Ayesha Ahmed (CamExam), Dr Therese Hopfenbeck (OUCEA), Carol Brown (OUCEA), Dr Victoria Elliott (University of York).



This project was initiated in response to the Secretary of State for Education's proposed reforms of the English examination system at age 16. In the [report](#), we summarised research relating to aspects of the reforms:

- How England competes in international tests
- The claim that there has been grade inflation
- The proposals to raise the level of challenge in examinations
- Familiarity with examination materials
- The desire to increase students' motivation
- The plan for students only to be tested at the end of the course, rather than within a modular system
- The proposal that the use of controlled assessment should be restricted
- The proposal that examinations should not be tiered

Assessment reform is an ongoing process in England, with many issues being revisited periodically. This report focuses on current proposed reforms, but we have framed it in a way that we hope will be useful for future reforms as well as for people in other countries. We report a great deal of relevant research that has been conducted on the reform topics, although the answers to some of our questions remain inconclusive.

### ***Rater Effects – use of multilevel modelling to monitor raters***

Funder: Pearson UK. Principal Investigator: Professor Jo-Anne Baird; Collaborators: Dr Daniel Caro (OUCEA), Malcolm Hayes (Pearson), Alex McKee (Pearson); Kath Thomas (Pearson), Ed Wolfe (Pearson), George Leckie (Bristol).

The Pearson rater effects project is investigating the value of multilevel modelling (MLM) in evaluating leniency/severity effects on the marking of students' work and providing feedback to marking monitoring systems. Examination data from seven question papers provided by Edexcel is being used for this purpose. This is part of a larger project, the Marking Quality Assurance Project, in which the applicability of different statistical techniques to study rater effects is being investigated.

### ***Investigation into the predictability of the Irish Leaving Certificate Examinations***

Funder: State Examinations Commission, Ireland; Principal Investigator: Professor Jo-Anne Baird; Collaborators: Professor Jannette Elwood (Queen's University, Belfast), Dr Therese Hopfenbeck (OUCEA), Dr Daniel Caro (OUCEA), Dr Ayesha Ahmed (CamExam), Rhian Barrance (Queen's University, Belfast), Natalie Usher (OUCEA), Christine Paget (OUCEA).

OUCEA has been commissioned by the State Examinations Commission in Ireland to investigate the [predictability of Irish Leaving Certificate examinations](#). A common theme for national assessment systems is the central question regarding whether students are being encouraged to learn the right material in the right ways; ways which will foster their future economic and social success. The research will be carried out in three phases: review of the literature; examination



materials analysis; and teachers and learners interviews and questionnaires in Ireland. A report will be submitted to the State Examinations Commission by the end of 2013.

### ***Causal comparability project***

Jo-Anne Baird, Daniel Caro and Lorena Ortega have begun working on a project that will address a conceptual debate in the literature about methods for linking examination standards. One approach is to compare standards using statistical models which use variables as controls for features of the (different) populations taking two examinations, such as those offered by different examining boards. Which variables should be entered as controls and whether they have the same effects upon each of the examinations has been controversial in the literature. In this project, we aim to demonstrate empirically some of the problems and to consider further whether the effects of the variables are direct or indirect.

### ***R package***

Daniel Caro has developed a statistical package in R. Its purpose is to analyse data from international assessments (PISA, TIMSS, and PIRLS) using the R language and environment, which is open source and freely available. Results in PISA and TIMSS/PIRLS international reports can be reproduced using the R package '[intsvy](#)'. The international assessment data cannot be analysed with standard software due to complexities associated with study designs (e.g. two-stage sampling and rotated test forms). Package 'intsvy' takes into account these complexities in the calculation of statistics and associated standard errors.

### ***Causal inference in educational studies with observational data***

Daniel Caro is working towards the preparation of research articles that investigate causation in educational research with observational data, that is, in survey and assessment studies with non-experimental designs. Causation is viewed from two perspectives: one from the econometric literature that focuses on causal effects, and another from the psychology and political science literature that is interested in the causal mechanisms. Daniel has presented a review of related statistical techniques at two Quantitative Special Interest Group sessions at the Department of Education.





## Publications in 2012-13

A number of articles were submitted for peer review during this period and the following were in print before the end of June 2013. Publications are available on the OUCEA website.

### *Guest-edited Special Issues*

- Baird, J. & Black, P. (2013) (Eds) *Research Papers in Education*, 28 (1), February 2013. Special Issue: [The reliability of public examinations](#).
- Oancea, A. and Hopfenbeck, T.N. (2012) (Eds) [International Datasets and Comparisons in Education](#), *Research Intelligence*, 119, British Educational Research Association.

### *Peer-reviewed journal articles*

- Baird, J. & Black, P. (2013) [Test theories, educational priorities and reliability of public examinations in England](#), *Research Papers in Education*, 28 (1), 5-21.
- Baird, J., Rose, J. & McWhirter, A. (2012) [So tell me what you want: a comparison of FE college and other post-16 students' aspirations](#), *Research in Post-Compulsory Education*, 17 (3), 293-310.
- Caro, D.H. & Cortés, D. (2012) [Measuring family socioeconomic status: An illustration using data from PIRLS 2006](#), *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 5, 9-33.
- Caro, D.H. & Lenkeit, J. (2012) [An analytical approach to study educational inequalities: 10 hypothesis tests in PIRLS 2006](#), *International Journal of Research and Method in Education*, 35 (1), 3-30.
- Caro, D.H. & Mirazchiyski, P. (2012) [Socioeconomic Gradients in Eastern European Countries: evidence from PIRLS 2006](#), *European Educational Research Journal*, 11 (1), 96-110.
- Caro, D.H. & Schulz, W. (2012) [Ten Hypotheses about Tolerance among Latin American Adolescents](#), *Citizenship, Social and Economics Education*, 11 (3), 213-234.
- Daly, A., Baird, J., Chamberlain, S. & Meadows, M. (2012) [Assessment reform: students' and teachers' responses to the introduction of stretch and challenge at A-level](#), *Curriculum Journal*, 23 (2), 173-187.
- Hopfenbeck, T.N., Throndsen, I., Lie, S. & Dale, E.L. (2012) [Assessment with distinctly defined criteria: A research study of a National project](#), *Policy Futures in Education*, 10 (4), 421-433.
- Simpson, L. & Baird, J. (2013) [Perceptions of trust in public examinations](#), *Oxford Review of Education*, 39 (1), 17-35.

### *Editorials*

- Baird, J. (2013) [The currency of assessments](#), *Assessment in Education: Principles, policy and practice*, 20 (2), 147-149.
- Baird, J. (2012) [Editorial](#), *Assessment in Education: Principles, policy and practice*, 19 (4), 389-391.
- Baird, J. (2012) [Do we need marking at all?](#), *Assessment in Education: Principles, policy and practice*, 19 (3), 277-279.

Baird, J. & Black, P. (2013) [The reliability of public examinations](#), *Research Papers in Education*, 28 (1), 1-4.

Hopfenbeck, T.N. (2012) [The Role and Value of International Datasets and Comparisons in Education Research](#), Guest Editorial in *International Datasets and Comparisons in Education*, *Research Intelligence*, 119, British Educational Research Association, 7-8.

### ***Edited books***

Hopfenbeck, Therese N.; Kjærnsli, Marit & Olsen, Rolf Vegar (red.) (2012). *Kvalitet i norsk skole: Internasjonale og nasjonale undersøkelser av læringsutbytte og undervisning*. Universitetsforlaget. [Quality in Norwegian Schools: international and national studies of learning and teaching.] ISBN 9788215020044. 296 s.

### ***Book chapters***

Caro, D.H. (2012) Evidencia causal en estudios educativos con bases de datos observables. [Causal evidence in educational studies with observational data]. In E. Vásquez (Ed) *Inversión social: indicadores, bases de datos e iniciativas*. [Social investment: indicators, data sets and initiatives]. Lima: Universidad del Pacífico.

Hopfenbeck, T.N. (2013) What did you learn in school today?, in: J. Hattie, T.S. Wille, M. Hermansen, T.N. Hopfenbeck, C. Madsen, P. Kirkegaard, H. Bjerresgaard, C.E. Weinstein, I. Bråten & R. Andreassen (Eds) *Feedback og vurdering for læring*, in Danish. ISBN: 978-87-7281-685-2.

Lillejord, S. & Hopfenbeck, T.N. (2013) Vurdering og læring i skolen, [Assessment and learning in school.] in: Lillejord, S., Manger, T. & Nordahl, T. (Eds) *Livet i skolen 2: Grunnbok i pedagogikk og elevkunnskap. Lærerprofesjonalitet*, 231-259.

### ***Book reviews***

Flórez, T. (2013) Book review of “Assessment reform in education: policy and practice”, *Assessment in Education: Principles, Policies and Practice*. Published [online](#) April 2013.

### ***Public reports***

Baird, J., Ahmed, A., Hopfenbeck, T., Brown, C. & Elliott, V. (2013) [Research evidence relating to proposals for reform of the GCSE](#). OUCEA Report.

Baird, J., Hayes, M., Johnson, R., Johnson, S. & Lamprianou, I. (2013) [Marker effects and examination reliability: A comparative exploration from the perspectives of generalizability theory, Rasch modelling and multilevel modelling](#). Report commissioned by the Office of Qualifications and Examinations Regulation. Ofqual/13/5261.

Olsen, R.V., Hopfenbeck, T.N., Lillejord, S. & Roe, A. (2012) [Elevenes læringssituasjon etter innføringen av ny reform](#), *Acta Didactica Oslo*. 1/2012.

## Research dissemination

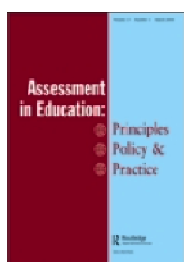
Staff and students have presented at the following external events.

Presenter(s)	Event	Date	Title
Baird, Meadows (AQA) and Leckie (University of Bristol)	Symposium on rater effects at the International Testing Conference, Amsterdam  Co-presenters of the symposium: Alvaro Arce (Pearson), Rense Lange (Illinois State Board of Education, Avi Allalouf, Galit Klapfer, Marina Fronton (National Institute for Testing and Evaluation, Jerusalem), George Engelhard and Stefanie Wind (Emory University, Iasonas Lamprianou (University of Cyprus)	5 Jul 2012	Monitoring raters: The effect of quality assurance system
Flórez	ECER Conference, University of Cadiz, Spain	18-21 Sep 2012	The Need for Complexity in Education Research: A Polysystemic Approach to Assessment for Learning Policymaking and Practice
Baird	Pearson conference, St. Anne's College, Oxford	20 Sep 2012	Standards at GCSE, as easy as EBC
Baird & Elwood	Ofqual Board Meeting	5 Oct 2012	Reform of the GCSE
Baird, Hopfenbeck, Flórez & Isaacs (Institute of Education)	13 <sup>th</sup> Annual AEA-Europe conference, Berlin	8-10 Nov 2012	Educational assessment policy reform: what can we learn across borders?
Flórez	13 <sup>th</sup> Annual AEA-Europe conference, Berlin	8-10 Nov 2012	Keynote: Systems, Ideologies and History: A Three-Dimensional Absence in the Study of Assessment Reform Processes
Baird	Domus Seminar, St. Anne's College, Oxford		Grade inflation: the qualifications arms race
Baird	UCAS conference, Nottingham	5 Feb 2013	Grade inflation in public examinations
Baird	SCORE conference, London	26 Feb 2013	Examination reform: return of the GCSE
Baird	CEMO, University of Oslo	27 Feb 2013	Grade inflation in public examinations
Flórez	International Seminar of the Finnish Doctoral Programme in Education and Learning (FiDPEL), 'Education Policy and Politics: Perspectives on Contemporary Developments', Oxford	9-11 Apr 2013	Discussant on Reijo Miettinen's lecture on Innovation and Human Capabilities in an Enabling Welfare State

Presenter(s)	Event	Date	Title
Hopfenbeck	Vasant Valley School, New Delhi, India	7 May 2013	Assessment for Learning – workshop for Learn Today team
Baird	St. Anne's College, Oxford	8 May 2013	Research evidence relating to proposals for reform of the GCSE
Hopfenbeck	Vasant Valley School, New Delhi, India	10 May 2013	Assessment and Self-Regulation – workshop for teachers
Flórez	Oxford Education Society and OUCEA seminar on 'Educational standards and qualification reform'. Department of Education	18 May 2013	Respondent
Hopfenbeck	Governing Complex Education Systems. OECD Third thematic Conference, Paris.	17-18 Jun 2013	<i>GCES case study Norway: Assessment for Learning</i>
Hopfenbeck	Transition conference in Ireland	24 Jun 2013	<i>Predictability in the Leaving Certificate?</i>

## OUCEA Events

Previous years' events have contributed to two Special Issues of journals in 2013.



Professor Richard Daugherty hosted an invited research seminar at St. Anne's College on 20<sup>th</sup> - 22<sup>nd</sup> June 2011. Professor Val Klenowski, one of the participants at the seminar has Guest Edited a Special Issue on Moderation Practice and Teacher Judgement. Presentations at the seminar formed the basis of a number of the articles in the [20 \(1\) edition](#) of *Assessment in Education: principles, policy & practice*.



Professor John Gardner convened the Public Understanding of Assessment seminar series at OUCEA in 2011-12. An associated Special Issue of the *Oxford Review of Education* was Guest Edited by Professor Gardner. The title of the Special Issue is [The Public Understanding of Assessment](#) – 39 (1).



Fourteen events were held this year, with approximately 290 students, academics and professionals attending in total. Some of the events were recorded and are available as [podcasts](#) on the OUCEA website.

### **Invited Symposium on A-level reform 4 September 2012**

The 2010 Schools White Paper on the Importance of Teaching announced that Higher Education should be more involved in A-level development and signalled a change to re-sitting arrangements. Subsequently, a number of policy announcements were made relating to the re-calibration of standards, Higher Education endorsement of A-levels and changes to the structure of A-levels. This symposium was a forum for the presentation and discussion of research relevant to the reforms. Participants were from Higher Education, industry and government and the main purpose of the symposium was to bring together research, practitioners and policy makers to critically analyse the factors that will affect the success of this reform.

The following presentations were made:

Susannah Wright – A-Levels and Higher Education: Academic and admissions staff views

Jo-Anne Baird – Assessment reform

Tina Isaacs and Alison Matthews – Lessons learned from C2K and the Diploma

Michelle Meadows – Re-setting the watch: recalibration and discrimination

Rose Clesham – An analysis of A level demand

The [report of the symposium](#) is available on the OUCEA website and it was submitted to Ofqual in response to their consultation on A-level reform.

### **Masterclass “Introduction to Diagnostic Modelling for Large Scale Assessment Data” 4-5 October 2012**



Dr Matthias von Davier led this workshop. The general diagnostic model (GDM) represents a framework for modelling item response data with multi-dimensional IRT-Models and discrete diagnostic classification models. A generalisation of the GDM integrates features of discrete mixture distribution models for item response data as well as multilevel diagnostic models. An introduction of this modeling framework was given that relates these recent developments to well-known psychometric model approaches such as the Rasch model,

item response theory, and latent structure analysis. The workshop gave participants an opportunity to analyse example datasets with the models presented over the course of the two days.

### **OUCEA Seminar – the philosophy of constructs 28 March 2013**

Dr Josh McGrane, University of Western Australia

### **Department of Education Advanced Course “An introduction to R: applied examples with international assessment data” 9 May 2013**

Daniel Caro led this workshop for staff and students of the Department of Education.

### **Invited Symposium on GCSE mathematics reform 26 October 2012**

This event was jointly held by the Advisory Committee on Mathematics Education (ACME) and OUCEA.

The symposium was a forum for the presentation and discussion of research and policy relevant to the proposed GCSE Mathematics reforms. These include five-year franchise arrangements for the delivery of GCSE Mathematics, the removal of modular assessment, tiering and calculators. Participants were from ACME, Higher Education, industry and government and the main purpose of the symposium was to bring together research, practitioners and policy makers to analyse critically the factors that will affect the success of this reform.

[Papers are available](#) on the OUCEA website. The following presentations were made:

Jennie Golding – Purposes and Principles: an ACME view  
Richard Browne – Exploring some implications of the proposals  
Josh Beattie – Key Stage 4 Qualification Reform  
Jo-Anne Baird – The qualifications market

**Teachers from Torstad skole in Norway** visited OUCEA on 12<sup>th</sup> April 2013 for a continuing professional development programme on the themes of assessment for learning and self-regulation. The programme was led by Dr Therese Hopfenbeck and Professor Gordon Stobart.



**Educational Standards and Qualification Reform** – An event organised in collaboration with the Oxford Education Society (OES) 18 May 2013

The [conference](#) involved the presentation and discussion of research, policy and practice relating to current reforms. These include five-year franchise arrangements for the delivery of GCSE, the removal of modular assessment, tiering and examination aids. At A-level, universities will be involved in the content of the curriculum and examinations, all examinations will be end-of-course assessments and there will be a re-calibration of outcome standards. The main purpose of the event was to bring together academics, practitioners and policy makers to critically analyse the factors that will affect the success of these reforms for improving young people's education.



The speakers and the titles of their presentations were:

- Glenys Stacey, Chief Executive of Ofqual: Principles of qualification regulation and reform
- Michelle Meadows, Director of Research and Policy, AQA: From benchmark to judge's bench: an insider's view of the causes of the GCSE English exams crisis
- Jo-Anne Baird, Pearson Professor of Educational Assessment, University of Oxford: Grade inflation
- Paul Newton, Professor of Educational Assessment, Institute of Education, University of London: Standards, transparency and reasonable expectations

### **Department of Education Public Seminars**

Professor Anil Kanjee, Tshwane University of Technology, South Africa – *Assessment and education quality in Sub-Saharan Africa: Prospects & Pitfalls*, 29 October 2012

Professor Paul Newton, Institute of Education, University of London – *Does it matter what 'validity' means?* 4 February 2013

Dr Therese Hopfenbeck, and Associate Professor Astrid Tolo, University of Bergen - *Assessment policy in Norway – an OECD case study* Respondents: Jenny Ozga and Solvi Lillejord, Research Council of Norway, 25 February 2013



## Department of Education Quantitative Methods Special Interest Group Seminars

Dr Daniel Caro, *Causal inference in educational studies with observational data – Part 1*, 21 January 2013.

Dr Jenny Lenkeit, *How effective are educational systems? A value-added approach to study trends in international assessment studies* (Amsterdam Centre for Inequality Studies, University of Amsterdam), 13 May 2013.

Dr Daniel Caro, *Causal inference in educational studies with observational data – Part 2*, 6 May 2013.

## Student Assessment Network

The [Student Assessment Network](#) (StAN) is a student-led network of graduate students who have an interest in assessment. StAN organises at least two meetings every term. One meeting hosts a speaker (an academic, policy maker, professional, etc. in assessment) and the other one gives the opportunity for students to present their work and get constructive feedback from peers and academics in the field of assessment. StAN also organises ad-hoc presentations, consultations, reading groups and writing groups.

Agni Paramita is the Co-ordinator of the network. Eight Student Assessment Network events have been held this year, details of which are below.

13th June 2013: Student Presentations. In these sessions students present their work and receive feedback from their peers – whatever the phase of their research.

30th May 2013: Writing Group. In these sessions students discuss and provide feedback on written pieces from two DPhil students in the write-up phase of their research projects. This is a great opportunity for students to engage with fellow researchers of assessment, and practise reading critically and providing feedback on writing.

28th March 2013: Dr Joshua McGrane, University of Western Australia, “A shaky discussion of the foundations of measurement (by a social scientist)”.

31st January 2013: Dr Daniel Caro, OUCEA Research Fellow, “Ten Hypotheses about Tolerance among Latin American Adolescents”.

25th February 2013: Associate Professor Astrid Tolo, University of Bergen, “Teacher Learning in Relation to Assessment”, with examples from the Norway experience.

21st November 2012: Student presentations from Carol Brown (Relationship between summative assessment and learning identity in A-level students in the UK), Agni Paramita (School effects on student achievement in Indonesia’s Junior Secondary education summative assessments: Do resources matter?) and Desmond Tan (Student teachers learning Assessment for Learning in schools)

25th October 2012: Keynote by María Teresa Flórez in preparation for AEA-Europe conference

18th October 2012: Professor John Gardner, University of Stirling, “The use of assessment to support learning”



StAN Reading Group



Carol Brown, doctoral student



María Teresa Flórez, doctoral student

## Committee Membership and Expert Groups

Therese Hopfenbeck was appointed to Chair the Norwegian Centre for Learning Environment and Behavioural Research in Education. Currently, our committee work outside the Department of Education includes: Cambridge Colloquium Steering Board, Ofqual's Standards Advisory Group, Oxford University's Admissions Executive Sub-Group on Testing, Oxford University Working Group on Classification, Pearson Assessment Expert Group – UK and the Scottish Qualifications Authority Qualifications Committee. We also provided consultancy to the National Council for Curriculum and Assessment in Ireland, on the grading structure of their examinations and Jo-Anne Baird was a member of the Dutch Programme Council for Educational Research Evaluation Committee on the Gifted and Talented Research Programme.

## Media

OUCEA now has a presence on both LinkedIn and Twitter. The LinkedIn group is called [Oxford University Centre for Educational Assessment](#) and the Twitter account is @Baird\_jo\_anne.

Our research was covered in the following media articles:

[TES](#) 27 April 2013

[Oxford University](#) 8 May 2013

[BBC News](#) 8 May 2013

[Yorkshire Post](#) 8 May 2013

[Guardian](#) 10 May 2013

[BBC Radio 5](#) 11 June 2013

[BBC Radio Derby](#) 11 June 2013

## Future Plans

Over the coming year, we will be involved in the organisation of several European conferences on assessment and plan to contribute to many academic events. We have plans to host academic visits from Professor David Andrich, Dr Michelle Meadows and Dr Edward Wolfe.

## Students

Offers have been made to doctoral students for entry in October 2013. This would bring the number of doctoral students currently studying at OUCEA to 10, which is a large critical mass of assessment students that will not be found in many institutions internationally.

## Research Programme

In addition to the doctoral research projects, our programme will include the following projects:

- Predictability of examinations, and its effects
- Grade inflation
- Governing Complex Education Systems
- Policy Effects of PISA
- Rater effects and quality of marking
- Comparability of standards
- Trust in examinations
- Developments of statistical packages to analyse test data



## Appendix: Abstracts

### Abstracts from journal articles published 1 July 2012 to 30 June 2013

**Baird, J. & Black, P. (2013) [Test theories, educational priorities and reliability of public examinations in England](#), *Research Papers in Education*, 28 (1), 5-21.**

Much has already been written on the controversies surrounding the use of different test theories in educational assessment. Other authors have noted the prevalence of classical test theory over item response theory in practice. This Special Issue draws together articles based upon work conducted on the Reliability Programme for England's examinations regulator, Office of Qualifications and Examinations Regulation. One strand of the work was methodological, in which we noted the advantages and assumptions of different approaches to investigating reliability of public examinations. A feature of the field that has not been well documented is why psychometrics in general does not always fit the educational, cultural and political priorities of public examinations. Psychometrics has implications for test design and reliability. Public examinations are: curriculum-embedded; it is desirable that the domain is transparent; there are many curriculum changes (which makes item banking more problematical and costly); curriculum exposure affects performance; public examination questions might need to be secure until the examination is released and then, made public; pre-testing is infrequent; there is non-random syllabus representation; item independence assumptions are often broken; score distribution might not be expected to be normal; complex assessment designs compose whole qualifications; a latent trait might not be assumed; unidimensionality might not be considered important; and there might be very weakly described constructs. Implications for measurement of reliability in this context in England include less concern for internal reliability or occasion-related factors and an emphasis upon standard-setting and inter-rater reliability. Psychometricians sometimes view this approach as technically weak and behind the times, but we raise the prospect that psychometrics could seem like an answer to somebody else's problems if the curriculum, questions and washback upon learning are the main concern of the assessment designers. We agree with other authors that the field is currently under-theorised and raise the prospect that future theory needs to take account of the educational context of public examinations.

**Baird, J., Rose, J. & McWhirter, A. (2012) [So tell me what you want: a comparison of FE college and other post-16 students' aspirations](#), *Research in Post-Compulsory Education*, 17 (3), 293-310.**

This study compares the aspirations of students in further education (FE) colleges and non-FE post-compulsory settings. Within the context of a survey about their education, 928 students across 26 institutions were asked what their 'hopes, goals and plans' were, with space for five open responses. Coding of responses was based upon self-determination theory. Career aspirations were most frequently mentioned, followed by educational aspirations (approximately one quarter of responses for each). Students from FE settings mentioned career aspirations more frequently and educational aspirations less frequently than students in non-FE settings. Aspirations such as personal happiness and life satisfaction were also mentioned by students. The need for FE colleges to support students in developing high aspirations was highlighted. Furthermore, a wider framework of student aspiration for use in both research and educational practice is needed.

**Caro, D.H. & Cortés, D. (2012) [Measuring family socioeconomic status: An illustration using data from PIRLS 2006](#), *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 5, 9-33.**

Many analyses of educational outcomes include a single socioeconomic status (SES) index as a predictor or for statistical control. In large-scale assessment research, these analyses have primarily



examined the influence of SES on academic achievement in schools and the influence of other individual, family, and school variables when controlling for SES. The findings have helped elucidate the mechanisms underlying the influence of SES and identify possible avenues for reducing it. The measurement of SES is critically important with respect to the findings and implications of these analyses. This paper demonstrates a step-by-step procedure for calculating a family SES index using data from the Progress in International Reading Literacy Study (PIRLS) 2006. Variables reflecting parental education, parental occupational status, and family wealth were reduced into a single SES index using principal component analysis (PCA). The conceptual definition of SES, its reliability and crossnational comparability are discussed, and recommendations for further research and survey developers are offered. The presented procedure can be extended to other large-scale studies of educational achievement. Its illustration is the main contribution of this paper.

**Caro, D.H. & Lenkeit, J. (2012) [An analytical approach to study educational inequalities: 10 hypothesis tests in PIRLS 2006](#), *International Journal of Research and Method in Education*, 35 (1), 3-30.**

The paper draws on the work of Willms [2006. *Learning divides: Ten policy questions about the performance and equity of schools and schooling systems*. Montreal: UNESCO Institute for Statistics] to present an analytical approach to the study of academic achievement disparities related to family socioeconomic status. The approach is illustrated by evaluating 10 hypotheses with two-level and three-level hierarchical linear models using data from the Progress in International Reading Literacy Study 2006. For each hypothesis, the underlying theory, statistical model, and critical model test are presented and the results are discussed. The analytical approach can be generalized to other studies and data sets. The results help to understand how inequalities are configured at the within-country and between-country levels.

**Caro, D.H. & Mirazchiyski, P. (2012) [Socioeconomic Gradients in Eastern European Countries: evidence from PIRLS 2006](#), *European Educational Research Journal*, 11 (1), 96-110.**

This article analyses educational inequalities related to socioeconomic status (SES) in 12 Eastern European countries that participated in the International Reading Literacy Study (PIRLS) 2006. Economies and educational systems of these countries have undergone critical transformations since the fall of communism. The authors' analyses, using data collected almost 20 years after this period, help explain how these transformations affected the equity and quality of educational outcomes in the region. For each country, overall inequalities as well as inequalities between schools and within schools are estimated with regression models and represented graphically with socioeconomic gradient lines. A possible trade-off between equity and quality of outcomes is explored, identifying countries that have been relatively successful at attaining both educational goals. The extent to which the school SES explains achievement gaps between rural and urban schools is analysed. The results point to country groupings that are reasonably consistent with regional classifications of educational systems postulated in the literature.

**Caro, D.H. & Schulz, W. (2012) [Ten Hypotheses about Tolerance among Latin American Adolescents](#), *Citizenship, Social and Economics Education*, 11 (3), 213-234.**

Discrimination on the grounds of sexual orientation, race, mental and/or physical disability, religious background, HIV/AIDS status, and ethnic origin affects the well-being of minorities and society in general. Recent research in North America underscores the importance of contact with diverse networks, intergroup discussions, a social dominance orientation, religious beliefs, and the school climate, among other factors, to explain tolerance for minorities. Theoretical and empirical work in Latin America is less extensive and has been limited by the lack of quantitative data. This article evaluates 10 different hypotheses about tolerance, using data from eighth-grade students in six Latin American countries that participated in the International Civic and Citizenship Education Study in

2009. Overall, the results provide support for most hypotheses. Notably, countries with relatively more positive views of minorities tend to exhibit very small differences in tolerance attitudes among students of varying socio-demographic characteristics and, apparently, also more effective mechanisms for promoting tolerance through schools.

**Daly, A., Baird, J., Chamberlain, S. & Meadows, M. (2012) [Assessment reform: students' and teachers' responses to the introduction of stretch and challenge at A-level](#), *Curriculum Journal*, 23 (2), 173-187.**

This paper describes an exploration into a reform of the A-level qualification in England in 2008; namely, the introduction of the 'stretch and challenge' policy. This policy was initiated by the exams regulator and determined that exam papers should be redesigned to encourage the application of higher order thinking skills, both in the classroom and in examinations. The present study comprised two strands that explored the perceptions of students ( $n = 39$ ) and teachers ( $n = 27$ ) regarding the degree to which the incorporation of opportunities for stretch and challenge in the new examination papers had been achieved, and the likely effects on teaching, learning and exam preparation. On the whole, students and teachers welcomed the stretch and challenge policy and there were some indications that changes to the design of question papers could have some positive backwash effects.

**Hopfenbeck, T.N., Throndsen, I., Lie, S. & Dale, E.L. (2012) [Assessment with distinctly defined criteria: A research study of a National project](#), *Policy Futures in Education*, 10 (4), 421-433.**

This article explores changes in teachers' beliefs and practice concerning assessment after participating in a project for improving assessment practices in Norwegian schools. The project was initiated by the Norwegian Directorate for Education and Training in 2008, and included a total of 77 schools, more than 600 teachers and a sample of their students. The main idea was to develop formative assessment practices in the classroom in the form of distinct criteria for clarification of how to reach curriculum goals. Furthermore, the project intended to increase students' motivation by using such formative assessment practices. The data presented in this article is from the research evaluation of the project. So far, the main challenge seems to be how to develop assessment criteria and how to use them in practice.

**Simpson, L. & Baird, J. (2013) [Perceptions of trust in public examinations](#), *Oxford Review of Education*, 39 (1), 17-35.**

Over recent years, the credibility of public examinations in England has increasingly come to the fore. Government agencies have invested time and money into researching public perceptions of the reliability and validity of examinations. Whilst such research overlaps into the conceptual domain of trust, trust in examinations remains an elusive concept. Little is known about what it means to trust in examinations and on what individuals base their trust judgments. This paper reports the findings of research into key stakeholders' perceptions of trust in the English A-level examination system through a series of focus groups comprising (separately) of examination board staff, students, higher education admissions tutors, teachers and the general public. Analysis of the data revealed a number of factors that are integral to the concept of trust in A-levels, including examination standards, the marking of examination papers, the syllabus, communication and provision of information, whether A-levels are fit for purpose and the mode of assessment. The fact stakeholders have different expectations and requirements of the A-level system, which are subject to change, means that it is difficult for the A-level system to engender the trust of all of its users all of the time.





© University of Oxford

Oxford University Centre for Educational Assessment (OUCEA)  
Department of Education  
15 Norham Gardens  
Oxford OX2 6PY  
UK

<http://oucea.education.ox.ac.uk>

