

| | | |
|---|----------------------------------|--|
| Institution: University of East Anglia | | |
| Unit of Assessment: 23 – Education | | |
| Title of case study: Improving the Quality, Reliability and Inclusivity of Large-scale Educational Assessments | | |
| Period when the underpinning research was undertaken: 2010 – 2018 | | |
| Details of staff conducting the underpinning research from the submitting unit: | | |
| Name(s): | Role(s) (e.g. job title): | Period(s) employed by submitting HEI: |
| Bryan Maddox | Associate Professor | Jan 2003 – to present |
| Period when the claimed impact occurred: 1 st August 2013 - 31 st July 2020 | | |
| Is this case study continued from a case study submitted in 2014? No | | |
| 1. Summary of the impact <p>International Large-Scale Assessments in education are now in use all over the globe, thanks in large part to the expansion of the Organisation of Economic Co-operation and Development (OECD) large-scale assessment programmes, and the impetus given to them by the UN Sustainable Development Goals. UEA research led by Dr Bryan Maddox, beginning in 2010 in the Mongolian Gobi Desert, has challenged deeply-held assumptions about the quality, reliability and inclusivity of International Large-Scale Assessments, in particular when applied across diverse populations, cultures, and contexts. Since then, Maddox has collaborated with global providers of large-scale assessments, from the OECD and national governments through to major bodies and corporations, and developed new analytical methods to identify, evidence, and resolve problems in current and future assessment designs. Across the period August 2013 to July 2020, these have helped improve the quality, reliability and inclusivity of assessments in more than 40 countries. Maddox's research led to a spin-out company in 2018 to further translate the research findings into practice internationally.</p> | | |
| 2. Underpinning research <p>The expansion of International Large-Scale Assessments in education began 20 years ago and has accelerated over the past decade. This is in part due to the globalisation of the OECD large-scale assessment programmes, such as PISA (the 'Programme for International Student Assessment') and PIAAC (the 'Programme for International Assessment of Adult Competencies'). The UN Sustainable Development Goals (Indicator 4.1.1) have accelerated this process, emphasising the importance of international measurement and comparison of educational outcomes, and impacting on assessment designs, national educational policies and curricula. However, concerns have been raised about the quality, reliability and inclusivity of these assessments when conducted among diverse populations.</p> <p>In 2010, the UNESCO Institute for Statistics commissioned Dr Maddox to investigate how culturally diverse populations respond to UNESCO's 'Literacy Assessment and Monitoring Programme' (LAMP). LAMP was developed to gather quality data on adult literacy and numeracy through new national household surveys. The LAMP instruments and methods were developed by and validated in six countries: El Salvador, Kenya, Mongolia, Morocco, Niger and Palestinian Autonomous Territories. Maddox was involved on site throughout the national assessment in Mongolia, conducting small-scale (micro-analytic) observations of face-to-face interaction in real-life educational assessments (R2). The findings of that study challenged the prevailing expert view about the conduct of test events, furnishing new evidence-based insights on assessment in practice in a developing country. The research with Mongolian nomadic herders showed how the cultural dynamics of the testing situation and other local contextual factors influence participation and response processes, reducing the quality and reliability of assessment data (R1,2,3). This was the first observational study to use techniques from anthropology to investigate International Large-Scale Assessments in a developing country.</p> | | |

Maddox was invited to collaborate with the OECD in evaluating their international large-scale assessments, PISA and PIAAC. PISA measures 15-year-olds' ability to use their reading, mathematics and science knowledge and skills to meet real-life challenges and has been applied in 72 countries since 2000. PIAAC assesses adults' proficiency in key information-processing skills - literacy, numeracy and problem solving - and gathers information and data on how adults use their skills at home, at work and in the wider community in over 35 countries. Maddox was commissioned to conduct fieldwork on their PIAAC assessment in Slovenia, and their 'PISA for Development' programme in Senegal. Those research projects uncovered further threats to assessment quality, reliability and inclusivity, through the interviewer-respondent interaction in the testing situation. They also revealed unexpected participant responses to test item content and a user interface that potentially compromised the test design (R4).

Maddox's underpinning research has involved methodological development, including the novel application of 'micro-analytic' observational research methods in assessment (R4). These comprised the use of digital techniques including video-ethnography (R4), and the use of eye tracking (R6) to observe assessment response processes, user experience and interaction. These methodological innovations in real-life test events created the opportunity to synchronise digital observational data with the log files that are routinely obtained in computer-based educational assessments, which then facilitated the investigation of data anomalies at scale (R4,5).

Maddox's dissemination and collaborations have been supported by an ESRC Seminar Series on 'International Large-Scale Assessments in Education' (R7), and a Leverhulme Trust International Academic Fellowship (R8).

3. References to the research

- R1 'Globalising Assessment: An Ethnography of Literacy Assessment, Camels and Fast Food in the Mongolian Gobi'.
Maddox, B.
Comparative Education, **2014**, 50(4), pp 474-489.
DOI: 10.1080/03050068.2013.871440
- R2 'An Anthropologist among the Psychometricians: Assessment Events, Ethnography and Differential Item Functioning in the Mongolian Gobi'.
Maddox, B., Zumbo, B.D., Tay-Lim, B. S-H., and Qu, D.
International Journal of Testing, **2015**, 15(4) pp 291-309.
DOI: 10.1080/15305058.2015.1017103
- R3 'Talk and Gesture as Process Data'.
Maddox, B.
Measurement: Interdisciplinary Research and Perspectives, **2017**, 15(3-4), pp 113-127.
DOI: 10.1080/15366367.2017.1392821
- R4 'Interviewer-Respondent Interaction and Rapport in PIAAC: The OECD Survey of Adult Skills',
Maddox, B.
Quality Assurance in Education. Special Issue on International Large-Scale Assessments, **2018**, 26(2), pp 182-195. DOI: 10.1108/QAE-05-2017-0022
- R5 'Investigating Testing Situations' in B. Maddox (Ed.) *International Large-Scale Assessments in Education*
Maddox, B., Keslair, F. and Javrh, P.
Bloomsbury Publishing, **2018**, Chapter 9, pp 157-172. ISBN: 9781350023628
- R6 'Observing response processes with eye tracking in international large-scale assessments: evidence from the OECD PIAAC assessment'.
Maddox, B., Bayliss, A., Fleming, P., Englehardt, P., Edwards, S.G. and Borgonovi, F.
European Journal of Psychology in Education, **2018**, 33, pp 543-558.
DOI: 10.1007/s10212-018-0380-2.

Grants

R7 Project: **'The Potentials, Politics and Practices of International Educational Assessments'**, (PI) Maddox, B.

Funder: ESRC.

Amount: GBP29,329. Dates: 2014 - 2016.

R8 Project: **International Academic Fellowship: 'Ecological Frameworks in Educational Testing'**, (PI) Maddox, B.

Funder: The Leverhulme Trust.

Amount: GBP6,105. Dates: 2014 – 2016.

4. Details of the impact

The research, developed through close collaboration with global and national organisations and with transnational commercial education providers, demonstrated for the first time that small-scale observational data on real-life assessment events could reliably be used to inform the analysis of test performance, response processes, user experiences, and interaction in large-scale assessments. This changed the mind-set of those commissioning, designing, delivering and evaluating international and national assessment programmes, leading to new approaches to test development and quality assurance, as detailed below.

Influencing changes in the assessment strategies of international organisations*Influence on UNESCO's assessment strategies*

The UNESCO Institute for Statistics recognised that the Mongolian research conducted by Maddox provided valuable insights into the conduct of large-scale, standardised educational assessments in culturally diverse settings (S1). A former Head of Learning Outcomes at the UNESCO Institute for Statistics (UIS) stated that:

"Dr Maddox's research findings suggested that ethnographic inquiry can shed light on different literacy issues, including the statistical measurement of these skills. It gave us some insights which then allowed us to identify some warning signals and make recommendations for a more cautious administration of LAMP. As such, Dr Maddox's work helped UIS reframe some of the instruments and tools used for LAMP. It also influenced UIS to recommend that participating countries consider conducting this sort of inquiry concurrently with LAMP implementation, as this would allow them to explore validity issues and the connection between literacy skills as portrayed by statistical methods and the uses of literacy in social life at a local level." (S1).

The findings have informed the UNESCO policy response to the Sustainable Goal 4 on Education (S2, S3). The UNESCO Institute for Lifelong Learning (UIL) Policy Brief on 'Direct assessments of adult skills and competencies' states:

"In some cultural environments, respondents may interpret and answer test questions in a way that could produce culturally biased results or counter-intuitive outcomes (Maddox et al., 2015). A strong infrastructure that meets generally accepted scientific standards is therefore needed to support and conduct data collection and analysis in the field (UIS, 2017). Broad-based tests that measure skills across cultures and social settings can be developed with appropriate investment and capacity-building efforts". (S3)

Influence on OECD's assessment strategies

The use of observational studies of assessment response processes and interaction gained institutional recognition by the OECD for the first time as a result of Maddox's research collaborations with them. This changed the OECD approach to the respondent/interviewer interaction that is at the heart of their global delivery of assessments. Senior Analyst for the 'PISA for Development' project (OECD) stated that:

"Bryan presented to the OECD team responsible for the project data from three assessment events with fine-grained evidence about PISA-D User Experience, and on assessment administration in a multi-lingual context. Bryan's findings were very helpful to the OECD and its

international contractors in fine-tuning the training of interviewers ahead of the PISA-D main survey data collection... The research had a significant impact on our preparation of interviewers prior to the Main survey data collection. Bryan's research highlighted the ways in which interviewers might influence the assessment if they were not properly trained". (S4).

Similar changes have been implemented for PIAAC, as their Senior Analyst reported:

"The observational studies of interview situations (in Slovenia) have stimulated reflection on respondent/interviewer interaction and, in particular, how to achieve an appropriate balance between the role interviewers can play in creating an environment in which test-takers are motivated to display their true ability and the possible biases that can be created by interviewers behaving in systematically different ways." (S5).

Taking account how features of the testing situation and test administration can affect participant experience and performance.

The adoption of the new techniques, processes and protocols has improved the quality, reliability, and inclusivity of large-scale, educational assessment data on children and adults in 39 countries (33 engaged with the second PIAAC cycle since the first cycle completed in 2017; Senegal, Cambodia, Paraguay, Panama, and Ecuador for PISA-D; and the national assessment systems of France [also in the second PIAAC Cycle], and Luxembourg) (S4, S5). The research for the OECD on the PIAAC assessment in Slovenia (R4) led to changes to the OECD assessment field protocols and quality assurance practice, which have been applied to improve the quality of assessment data in the second cycle of PIAAC currently taking place in 33 participating countries (S4, S5). Referring to Maddox's research in collaboration with the OECD (R4), the Senior Analyst at OECD (who is in charge of PIAAC) stated that:

"This work was commissioned with the aim of better understanding the nature of the testing situation, and the interactions that take place between respondents and interviewers, and respondents and the testing applications. The interest of this is to improve interview protocols and training and therefore reduce possible biases associated with variation in interviewer behaviour" (S5).

Improving assessment design, administration, data collection and analysis, and user experience

Maddox's research has led to the first systematic adoption of observational techniques, such as eye tracking, video-ethnography, and the analysis of talk and gesture, to improve assessment design, administration, and the use of assessment data, in the context of real-life programmes of large-scale educational assessment. Maddox presented research findings to the OECD and their technical partners (Microsoft, ETS, West and Capstan), leading OECD to change field protocols for the assessment of out-of-school youths and interpretation of assessment data, and informing OECD design of user interfaces to assist participants' use of the tablets provided for testing. Senior Analyst, PISA for Development (OECD) commented that:

"[Dr Maddox's] research also helped us understand the respondents' interface with the tablet computer used for the assessment" (S4).

PIAAC Analysts at the OECD have used Maddox's eye-tracking research when interpreting data collected during test participation. It helped them reassess the significance of periods of apparent inaction by test participants during testing, and to recognise that participants' offline activity must be considered when interpreting log-file data. A PIAAC Senior Analyst reported:

"The eye-tracking work has been important for the interpretation of log-file data (the electronic traces of test-taker/computer interaction). In particular, it has helped us to demonstrate that care needs to be taken with the interpretation of episodes of inaction. The absence of interaction with the testing application cannot be interpreted as a sign of cognitive inaction. Many actions by the respondent in the course of a test do not involve interaction with the test application – e.g., doing calculations with pencil and paper." (S5)

Developing new products and processes to enhance educational assessments

Maddox launched the spinout company Assessment Micro-Analytics Ltd in October 2018. The establishment of this company created a commercial basis for the routine application of micro-analytic methods in large-scale assessments. This led to government contracts covering entire national educational systems (France and Luxembourg), and contracts with the major commercial and non-governmental testing organisations, including Pearson UK, the Assessments and Qualifications Alliance (who have also invited Maddox to sit on their Research Committee), and American College Testing (where Maddox was a Visiting Scholar).

The work of Assessment Micro-Analytics Ltd with the Ministry of Education in France on national level computer-based assessments of secondary school mathematics led to the routine adoption of new observational techniques and processes; the re-design of test items to improve user experience; and generated insights into large-scale data on student achievement across the entire French secondary education system. The Deputy Director for Evaluation, Directorate for Evaluation, Performance and Prospective Studies (DEPP) of the Ministry of Education, France reported:

“Dr. Bryan Maddox's study contributes to the improvement of the national system of standardised assessments. Indeed, the Ministry is investing a lot in digital innovative formats for student assessment. Such innovative items involve numerous interactions and generate a significant amount of data. This new approach takes advantage of the benefits provided by a digital environment. It should make it possible to better assess pupils' skills in mathematics and to identify specific student profiles, in relation to the difficulties they encounter in mathematics. However, these new modalities raise important validity questions. Dr. Bryan Maddox's study, through a UX [user experience] approach, allows to question in a very fine detail the validity of these new components, both in their configuration and in the interpretation of the collected data. The stakes are high because these technologically-enhanced items are intended, in the future, to feed into the nation-wide assessments which take place each year on computers, and which concern all grade 6 and grade 10 students (1.6 million students each year). In addition, these items may also be used for formative assessment for pupils in the classroom and may be made more widely available in France, as well as in other countries.” (S6)

5. Sources to corroborate the impact

- S1 Testimonial from former Head of Learning Outcomes at the UNESCO UIS, 11.11.2020
- S2 UNESCO (2017) 'Implementation in Diverse Settings of the Literacy Assessment and Monitoring Programme (LAMP) Lessons for Sustainable Development Goal 4 (SDG 4)', UNESCO Institute for Statistics, Montreal, p. 51, 54.
- S3 UNESCO UIL (2020). 'Policy Brief 12: Direct Assessments of Adult Skills and Competencies', p.2.
- S4 Testimonial from Senior Analyst, PISA for Development Project (OECD), 02.11.2020
- S5 Testimonial from Senior Analyst, PIAAC (OECD), 12.11.2020
- S6 Testimonial from The Deputy Director for Evaluation, Directorate for Evaluation, Performance and Prospective Studies (DEPP) of the Ministry of Education, France, 16.12.2020.