UNDERSTANDING WHAT WORKS IN ORAL READING ASSESSMENTS

UNDERSTANDING WHAT WORKS IN ORAL READING ASSESSMENTS

UNESCO

The constitution of the United Nations Educational, Scientific and Cultural Organization (UNESCO) was adopted by 20 countries at the London Conference in November 1945 and entered into effect on 4 November 1946. The Organization currently has 195 Member States and 10 Associate Members.

The main objective of UNESCO is to contribute to peace and security in the world by promoting collaboration among nations through education, science, culture and communication in order to foster universal respect for justice, the rule of law, and the human rights and fundamental freedoms that are affirmed for the peoples of the world, without distinction of race, sex, language or religion, by the Charter of the United Nations.

To fulfil its mandate, UNESCO performs five principal functions: 1) prospective studies on education, science, culture and communication for tomorrow's world; 2) the advancement, transfer and sharing of knowledge through research, training and teaching activities; 3) standard-setting actions for the preparation and adoption of internal instruments and statutory recommendations; 4) expertise through technical co-operation to Member States for their development policies and projects; and 5) the exchange of specialised information.

UNESCO is headquartered in Paris, France.

UNESCO Institute for Statistics

The UNESCO Institute for Statistics (UIS) is the statistical office of UNESCO and is the UN depository for global statistics in the fields of education, science and technology, culture and communication.

The UIS was established in 1999. It was created to improve UNESCO's statistical programme and to develop and deliver the timely, accurate and policy-relevant statistics needed in today's increasingly complex and rapidly changing social, political and economic environments.

The UIS is based in Montreal, Canada.

COVER PHOTOS: Left and bottom right, © Dana Schmidt/The William and Flora Hewlett Foundation; top right, © Margarita Montealegre, Nicaragua; centre, © Uwezo, Kenya

BACK COVER PHOTOS: Top, © Margarita Montealegre, Nicaragua; bottom © Dana Schmidt/The William and Flora Hewlett Foundation

The UNESCO Institute for Statistics (UIS) led a collaborative project to formulate recommendations to guide practitioners when selecting, conducting and using oral reading assessments. The aim is to highlight basic principles that should be applied in the different stages of oral reading assessments—from planning and design to implementation and use of the resulting data. The recommendations are drawn from a collection of articles, which can be found online in the ebook, *Understanding What Works in Oral Reading Assessments,* at http://www.uis.unesco.org

Contributors

Organization	Author
Australian Council for Educational Research (ACER)	Marion Meiers
	Juliette Mendelovits
ASER Centre, Pratham India	Rukmini Banerji
	Shaher Banu Vagh
	Savitri Bobde
ASER Pakistan	Sehar Saeed
Concern Worldwide	Karyn Beattie
	Aine Magee
	Homayoon Shirzad
Concern Worldwide and University College Dublin	Jenny Hobbs
Creative Associates	Joy du Plessis
	Fathi El-Ashry
	Karen Tietjen
Durham University	Christine Merrell
	Peter Tymms
Education Development Center	Nancy Clark-Chiarelli
	Nathalie Louge
Instituto para el Desarrollo de la Democracia (IPADE)	Vanessa Castro Cardenal
Juarez and Associates, USAID Lifelong Learning Project	Cristina Perdomo
	Ana Lucía Morales Sierra
	Leslie Rosales de Véliz
	Fernando Rubio
Laboratoire de recherche sur les transformations économiques et sociales (LARTES), Jàngandoo	Binta Aw Sall
	Abdou Aziz Mbodj
	Diéry Ba
	Same Bousso
	Meissa Bèye
	Diadji Niang

Ministry of Education, Guatemala	María José del Valle Catalán
Ministry of Basic and Secondary Education, The Gambia	Momodou Jeng
RTI International	Keely Alexander
	Margaret M. Dubeck
	Amber Gove
	Emily Kochetkova
Save the Children	Ivelina Borisova
	Amy Jo Dowd
	Elliott W. Friedlander
	Lauren Pisani
Twaweza East Africa	Izel Jepchirchir Kipruto
	John Kabutha Mugo
	Mary Goretti Nakabugo
	Lydia Nakhone Nakhone
UNICEF	Manuel Cardoso
University College London, Institute of Education	Monazza Aslam
University of British Colombia	Linda Siegel
University of Oregon	Sylvia Linan-Thompson
University of Oxford	Pei-tseng Jenny Hsieh
The William and Flora Hewlett Foundation	Patricia Scheid
	Dana Schmidt
Women Educational Researchers of Kenya	Joyce Kinyanjui

Table of contents

Assessing Young Children: Problems and Solutions

CHRISTINE MERRELL, PETER TYMMS
Durham University

ABBREVIATIONS

app	Application
iPIPS	International Performance Indicators in Primary School
OECD	Organisation for Economic Co-operation and Development
PIPS	Performance Indicators in Primary Schools

1.   INTRODUCTION

Assessing young children around the age of starting school or earlier presents considerable difficulties and more arise if comparisons are to be made across different cultures and contexts. This article begins by expanding on some of those issues. We then proceed to suggest solutions to the apparently formidable difficulties. These solutions are based on the experience with the Performance Indicators in Primary Schools (PIPS) assessment—an assessment for use with children at the start of school. The authors have gained over 20 years of experience developing, adapting and successfully using the PIPS assessment in several countries. The PIPS Baseline Assessment was originally designed for formative use within schools (see Tymms, 1999; Tymms and Albone, 2002 for examples). It has subsequently been expanded to an international project for the study of children starting school and the progress that they make during their first school year in different parts of the world (www.ipips.org).

2.   CHALLENGES OF DESIGNING RELIABLE AND VALID ASSESSMENTS OF YOUNG CHILDREN'S COGNITIVE DEVELOPMENT

There are different approaches to assessing cognitive development, including posing questions, either in written format or verbally, asking children to perform a practical activity (possibly an open-ended investigation) and observing their responses or the way that they work and interact within an educational setting. Each of these methods is associated with a range of issues, some of which are common to all and others that are more assessment-specific. For example, we know from practice and research that:

- Many young children cannot read when they start school (Merrell and Tymms, 2007) and therefore traditional group assessments with administration instructions that require a certain level of reading ability are not feasible.

- Group assessments, such as a pencil-and-paper test, require an element of self-management and young children tend not to have the capacity to cope in such situations.

- Young children generally have a limited concentration span (Sperlich et al., 2015) and, consequently, the length of time that an assessment should take needs to be correspondingly short. This becomes an issue particularly if the method of assessment requires

a child to focus on answering questions or completing a directed activity.

- Young children have limited short-term memory capacity. Whilst an adult can be expected to hold seven novel pieces of information plus or minus two for a short time, a young child might only be able to concurrently hold two or three pieces of information (Demetriou et al., 2015). This means that obtaining reliable information from complex questions is not feasible.

- There are significant differences in the developmental levels within an age cohort. For instance, among children who are aged 4 years old in affluent countries, some will be reading fluently, doing complex sums and have an extensive vocabulary while others have not realised that text on a page is a code that carries meaning let alone possess the ability to identify letters (Tymms et al., 2014; Wildy and Styles, 2008a, 2008b; Merrell and Tymms, 2007). The latter group are probably unable to perform simple counting, don't recognise any digits and possess a vocabulary that may be extremely limited.

- If an assessment of children's cognitive ability is made entirely on the basis of observations, it is possible that they may fail to display their full potential. For example, a child may have an advanced understanding of mathematical concepts but if the activities in the setting do not challenge them to display this understanding, it will be missed. A further issue with observations is bias against individuals and groups (Harlen 2004, 2005; Sonuga-Barke et al., 1993; Wilmut, 2005).

- In addition to the challenges outlined above, consideration needs to be given to what should be assessed—the official curriculum of the country, variables that predict later success/difficulties or skills that are most malleable at the age of assessment? (Thompson and Nelson, 2001).

3. CHALLENGES OF DESIGNING ASSESSMENTS OF NON-COGNITIVE DEVELOPMENT IN YOUNG CHILDREN

The term 'non-cognitive skills' describes a collection of attributes and traits that represent the ways in which we think, our feelings, emotions and behaviour (Borghans et al., 2008). Non-cognitive skills continue to develop throughout our lives (Bloom, 1964). They include critical thinking and problem-solving skills, persistence, creativity and self-control. A recent report by the Organisation for Economic Co-operation and Development (OECD, 2015) emphasises the importance of non-cognitive skills for positive outcomes in life, using the big five personality dimensions that are widely recognised in psychology (openness, conscientiousness, extraversion, agreeableness and neuroticism) as an organizing framework. There are many descriptions of the big five traits, for example: Openness has been defined by Costa and McRae (1992) as the degree to which an individual is open to fantasies, aesthetics, feelings, new ideas and experiences. Conscientiousness was defined by Trapman et al. (2007) as the degree of dependability, organizational ability and degree to which an individual persists to achieve a goal. Extraversion is defined by Trapman et al. (2007) as the quantity and intensity of interpersonal interaction. Agreeableness is associated with being flexible in behaviour, cooperative and tolerant (Trapman et al., 2007). Neuroticism is described by adjectives such as anxious, touchy, nervous and unstable (Costa and McRae, 1992).

It cannot be assumed that young children possess a conceptual understanding of these attributes and traits or the ability to evaluate their own behaviours and actions in an objective way through an assessment. Their vocabulary is emerging and unlikely to be sufficiently sophisticated to be able to understand what is being asked and to be able to respond appropriately. Indeed, Soto et al. (2011) suggested that self-report questionnaires are only appropriate for children aged 10 years and over.

On the basis of research such as that of Soto et al. (2011), we must rely on adults who know the children to assess these non-cognitive areas on the basis of their knowledge built up through observations and interactions. But relying on adults to conduct these assessments has its challenges. Large classes can mean that teachers may not know each child well enough to provide a proper assessment. Most parents know their children well but some have low levels of literacy, making written surveys unreliable while interviews either in person or by phone are expensive to conduct. A further complication is that assessors interpret questionnaire statements in different ways.

4. CHALLENGES OF INTERNATIONAL COMPARISONS

Additional challenges are faced when making international comparisons of cognitive development and non-cognitive skills. Different interpretations of constructs and items arise across cultures as well as within cultures. For example, in a study by Merrell et al. (2013), teachers in Australia, England and Scotland were markedly different in their ratings of children's inattentive, hyperactive and impulsive behaviour. All international studies face issues with adaptations from one language to another given the subtle nuances conveyed through language. In reading assessments, there is the added challenge of working with different writing systems that could involve a major distinction between the alphabetic writing systems and the logographic writing systems used in some parts of Asia.

Ratings of non-cognitive skills are likely to be influenced by prevailing norms of behaviour and by individual perceptions (e.g. Merrell et al., 2013) and by variation in the standards applied by the assessors (Hosterman, 2009; Duckworth and Yeager, 2015).

For an international comparative study of children in their first school year, the varying ages at which they start school throughout the world needs to be taken into consideration because development in the early years varies greatly (Tymms et al., 2014);

one year at this stage can be a quarter of a child's life. In England, for example, the mean age at the start of school is 4.5 years and children may start school just after their fourth birthday. By contrast, children in Russia commonly start school at the age of seven (Kardanova et al., 2014). Family and cultural expectations will influence development as will a country's policy on early development. Can international comparisons of children's development and progress in their first year of school be valid? Can it yield useful information for policy and practice as well as increase our understanding of child development in general? In the next parts of this article, we identify ways to move forward in the face of the challenges arising from the issues discussed thus far.

5. ASSESSMENT PURPOSE: USING RESULTS FOR FORMATIVE, RESEARCH OR ACCOUNTABILITY

In an ideal world, we would want an assessment that provides information that is useful for the teacher, for national statistics and for research purposes. While assessment information is certainly useful for research and national statistics, when assessments become means of public accountability, they lose their formative purpose. In the case of the PIPS, the assessments are primarily intended to provide formative information for teachers. In order to ensure large scale use of the PIPS without harmful impacts, we create agreements that limit its use to formative purposes. We typically achieve this by creating an agreement on how the information is to be used that emphasises confidentiality. Results from the PIPS assessment are fed back to schools via a secure website where each participating school can see only their own results. The reports are a combination of charts and tables with both norm referenced and raw scores. The teacher can use the norm referenced scores to compare the development of their pupils with a representative sample. The raw scores provide detailed information about which questions each child answered correctly and incorrectly, revealing strengths and areas for development.

We do allow for the data to be used for research that aims to inform wider practice and policy but not for accountability. We formally agree with all stakeholders that the data is confidential—pupils, teachers and schools will not be identified when analysed and published for research purposes. We go so far as to say, in some areas, that if a school publishes their own data, they will be dropped from the project. This apparently aggressive stance provides schools with a reason not to disclose their results and yield to the pressures exerted from the public, journalists or higher authorities.

6. CONTENT OF THE ASSESSMENT

If the initial information from an assessment is to guide the teacher—as has been our intention when developing the PIPS Baseline Assessment—the assessment needs to include content that can provide comprehensive, reliable and valid information on what children know and the skills they possess on the road to becoming literate and numerate (Tymms et al., 2009). Depending on the purpose of the assessment, the results from the assessment should indicate each child's zone of proximal development so that the teacher can plan tailored learning experiences for specific children. If a sample of the class is assessed, the results will give the teacher an idea of its general ability level and its variation, which has some use but is more limited than information on all individuals. There are compromises to be made between collecting detailed information on all children in a large class and the time that it takes to accomplish this.

Information from an assessment of children's cognitive ability, such as the PIPS assessment, can be used as a predictor of later success or difficulties, and it can also be interpreted as an outcome measure for the time prior to assessment. It could be used as an outcome measure with which to evaluate early childhood development policy.

We have not focused on developing content specific to a country's official curriculum as many children at the start of school will not have followed an official curriculum yet. Curricula in early years tend to focus on developing general skills, such as personal and social development, basic language and precursors to reading, numeracy and motor skills rather than specific areas of learning, such as a specific period in history.

7. DESIGNING ASSESSMENTS FOR USE WITH CHILDREN DURING THEIR FIRST YEAR OF SCHOOL

Foremost, it should be noted that due to the stage of development among children in their first year of school, any assessment of a young child's cognitive development that is conducted before they start school or during their first school year must be conducted on a one-to-one basis with the assessor and the child if high-quality information is to be obtained. For the same reasons, the assessment must be completed within 15 to 30 minutes. Beyond this time, the validity of the data collected will drop as the children tire and their concentration ebbs. The administration costs will also rise if trained assessors are used to collect data for a research project.

The assessment must be robust so that it can produce reliable and valid results independently of the administrator, otherwise it is subject to the potential bias of that administrator (see Harlen, 2004, 2005 for examples of how assessments can be prone to bias). This is important if the results from assessments conducted by different assessors in different settings are to be meaningfully compared. Specialist knowledge and training should not be a necessary pre-requisite to obtaining high-quality information otherwise this limits the use of the assessment. The content of the assessment must be appropriate for children of a wide range of abilities within an early year's cohort. If the format and content of the assessment is to be appropriate for a wide range of ability and yet be administered within the time limit suggested earlier, then the only way to achieve this is to use an adaptive approach. If a child does not answer easy items correctly, they are not presented with more difficult ones but if a child is moving rapidly through the assessment and answering questions correctly, she/he is rapidly moved forward to more difficult

content. This approach not only addresses the range of abilities but it also decreases the time spent on the assessment. Furthermore, questions that are too difficult are not administered, reducing the probability of demoralising children with lower ability.

Can we envision a one-to-one adaptive assessment that can be carried out with the assessor using pencil and paper? Our own experience indicates that the assessors do not always follow the rules. An alternative method would be to use laptop computers but laptops can be expensive and security can sometimes be an issue when working in areas such as the favelas in Brazil or the townships of South Africa. We need an intelligent device that is inexpensive and widely used, and on which the assessment itself can be easily deployed. We found that a smartphone or tablet alongside a booklet provides the answer. The child and the assessor look at the booklet together. An application (app) is accessed by the assessor through the smartphone or tablet. The app selects items and the assessor records the child's responses electronically. The app contains rules that govern which questions are presented to the child on the basis of their answers. This relieves the assessor from having to follow adaptive rules and deciding which question to present to the child, which means that they can focus more on the child. We have used this approach successfully in the Western Cape of South Africa and in Russia. The assessment is divided into sections such as vocabulary, concepts about print, letter recognition and so on. Within each section, questions are organized in order of increasing difficulty in a series of sequences with stopping rules. Children start with easy items and if they answer questions correctly, they are presented with more difficult ones until they make a certain number of mistakes. The assessment then moves on to another sequence. This may be a more advanced section—for example, letter recognition if a child has demonstrated competence in concepts about print or a simple section of a different area of development such as counting. We have found that this approach can generate high quality data in just 15 to 20 minutes about children's early language, reading and mathematics development while

allowing the precocious few to demonstrate the full extent of their ability.

For assessments of non-cognitive development, we believe that we need to work with teachers rather than parents because their professional judgements are based on a wide experience of children of a similar age. For large classes, we suggest sampling because of the daunting workload that an assessment of every child in a class of 50 pupils would imply. However, the purpose of the assessment should guide the decisions on parameters. If the purpose is to inform policy, sampling children within classes and schools will provide a picture of children's development. If it is to inform classroom instruction, a sample will provide teachers with a profile of their class' abilities. If individual children with particular needs are to be identified, all children in the class should be assessed.

We have recently been developing a method to make comparable observations of children's behaviours from different contexts and cultures. It is debated if this can be achieved but we believe that it can be using short video clips of children (suitably anonymised) exhibiting different levels of a particular behaviour, such as attention, in a range of contexts. Teachers are asked to make their own ratings of the children in the clips and then to rate the children in their class. By analysing each teacher's ratings against the clips, we have evidence that we can establish the reliability, validity and severity of their scores and even evaluate if the construct being addressed is meaningful in different cultures.

8. COMPARISONS ACROSS COUNTRIES

Based on our experience, we found that comparisons across countries of some variables are virtually impossible while other variables lend themselves more easily to comparisons. The chart in **Figure 1** details this hypothesis, suggested by the authors.

Based on our research using the PIPS assessment, we found that some areas of development and

Figure 1. Comparison hypothesis

READING

MATHS

RHYME

HEIGHT

impossible ← → easy

SPECIFIC ACTIONS

CONCENTRATION

INDEPENDANCE
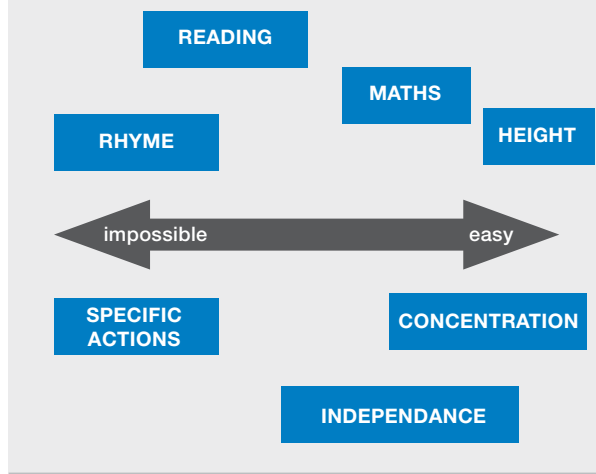
skills can be easily adapted across languages and cultures with few problems (e.g. simple arithmetic). Other areas cannot be adapted and compared so easily, such as the ability to spot rhyming words. In this case, words that rhyme in one language would not necessarily rhyme when translated into a different language. Alternative words would need to be used and this would change the level of difficulty of the items. It may be possible to devise an assessment of nonsense sounds and words which rhyme and would be unfamiliar to children in different countries but it is questionable whether valid data could be collected using this approach. There are certain behaviours that are particular to certain cultures, such as the use of head movements (for example, in the Indian sub-continent a head movement which means "yes" is seen in the west as meaning "no"). However, there are some behaviours and aspects of personality that can be compared— for example, conscientiousness, curiosity or the ability to empathise. Others may be linked to cultural norms. For example, it may be acceptable for a child to question a request made of them by an adult in one culture, indeed valued as a mark of curiosity or independence, but the same behaviour would be considered unacceptable in another and therefore not displayed. Caution is needed when interpreting the behaviour of children from different cultures.

In summary, some areas of development and skills can be compared across all cultures, others can

be compared across some cultures and some are unique to particular situations. All of these ideas are being put into practice within the International Performance Indicators in Primary School (iPIPS) project. Currently, data are being collected and analysed from South Africa, Russia, China, Brazil, England and Scotland.

CONCLUSION

We believe that despite the daunting challenges faced at the outset, with the PIPS and iPIPS we have developed an assessment system that works in different cultures and contexts. We are able to collect reliable and valid data in a short time with young children at varying developmental stages, which is useful to schools and at the same time can provide analyses for policymakers. Very similar assessments can be used with suitable adaptations across cultures and costs can be kept down so that work can be carried out effectively and efficiently.

REFERENCES

Bloom, B.S. (1964). *Stability and Change in Human Characteristics*. New York: John Wiley & Sons.

Borghans, L., Duckworth, A.L., Heckman J.J. and Bas ter Weel. (2008). "The Economics and Psychology of Personality Traits". *Journal of Human Resources,* Vol. 43, No. 4, pp. 972-1059.

Costa, P. T. Jr., and McRae, R. R. (1992). Revised *NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, Florida: Psychological Assessment Resources, Inc.

Duckworth, A.L. and Yeager, D.S. (2015). "Measurement matters: Assessing attributes other than cognitive ability". *Educational Researcher,* Vol. 44, pp. 237-251.

Demetriou, A., Spanoudis, G. and Shayer, M. (2015). "Mapping Mind-Brain Development". Farisco M. and Evers K. (eds.), *Neurotechnology and direct brain communication*. London: Routledge.

Kardanova E., Ivanova A., Merrell C., Hawker D. and Tymms P. (2014). *The role of the iPIPS assessment in providing high-quality value-added information on school and system effectiveness within and between countries*. Basic Research Program Working Papers. Moscow Higher School of Economics.

Harlen, W. (2004). "A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes". *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Harlen, W. (2005). "Trusting teachers' judgement: research evidence of reliability and validity of teachers' assessment used for summative purposes". *Research Papers in Education*, Vol. 20, pp. 245-270.

Hosterman, S.J. (2009). Halo Effects and Accuracy in Teacher Ratings of ADHD Symptoms: Influence of Ethnicity and Developmental Level. Ph.D. thesis, Lehigh University, USA.

Merrell, C. and Tymms, P. (2007). "What Children Know And Can Do When They Start School And How This Varies Between Countries". *Journal of Early Childhood Research*, Vol. 5, No. 2, pp.115-134.

Merrell, C., Styles, I., Jones, P., Tymms, P. and Wildy H. (2013). "Cross-country Comparisons of Inattentive, Hyperactive and Impulsive Behaviour in School-Based Samples of Young Children". *International Research in Early Childhood Education,*Vol. 4, No. 1, pp 1-17.

OECD (2015). *Skills for Social Progress: The Power of Social and Emotional Skills*. OECD Skills Studies. Paris: OECD Publishing. http://dx.doi.org/10.1787/9789264226159-en

Sonuga-Barke, E. J. S., Minocha, K., Taylor, E.A. and Sandberg, S. (1993). "Inter-ethnic bias in teachers' ratings of childhood hyperactivity". *Journal of Developmental Psychology*, Vol. 11, pp. 187-200.

Soto, C.J., John, O.P., Gosling S.D. and Potter J. (2011). "Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample". *Journal of Personality and Social Psychology*, Vol. 100, No. 2, pp. 330-348.

Sperlich, A., Schad, D. J., & Laubrock, J. (2015). "When preview information starts to matter: Development of the perceptual span in German beginning readers". *Journal of Cognitive Psychology*, *27*(5), 511-530.

Thompson, R. A. and Nelson, C. A. (2001). "Developmental science and the media: Early brain development". *American Psychologist*, Vol. 56, No. 1, p. 5.

Trapmann, S., Hell, B., Hirn, J.W., Schuler, H. (2007) "Meta-Analysis of the Relationship Between the Big Five and Academic Success at University". *Journal of Psychology*, 215, 132—151.

Tymms, P. (1999). *Baseline Assessment and Monitoring in Primary Schools: Achievements, Attitudes and Value-added Indicators*. London: David Fulton Publishers.

Tymms, P., Jones, P., Albone, S. and Henderson, B. (2009). "The first seven years at school". *Educational Assessment and Evaluation Accountability*, Vol. 21, pp. 67-80.

Tymms, P. and S. Albone (2002). "Performance Indicators in Primary Schools". A.J. Visscher and R. Coe. (eds.), *School Improvement Through Performance Feedback*. Lisse/Abingdon/Exton PA/ Tokyo: Swetz & Zeitlinger, pp. 191-218.

Tymms, P., Merrell, C., Hawker, D. and Nicholson, F. (2014). *Performance Indicators in Primary Schools: A comparison of performance on entry to school and the progress made in the first year in England and four other jurisdictions.* Department for Education: London. https://www.gov.uk/government/publications/performance-indicators-in-primary-schools

Wildy, H., & Styles, I. (2008). Measuring what students entering school know and can do: PIPS Australia 2006-2007. *Australian Journal of Early Childhood*, Vol. 33, No. 4, pp. 43-52.

Wildy, H., and Styles, I. (2008b). "What Australian students entering primary school know and can do". *Journal of Australian Research in Early Childhood Education*, Vol. 15, No. 2, pp. 75-85.

Wilmut, J. (2005). *Experiences of summative teacher assessment in the UK.* Qualifications and Curriculum Authority. London: QCA.