Measuring children's learning using tests from the Woodcock Johnson III

The central purpose of early childhood education is to advance children's learning and development. This helps to provide them the best start for life and prepare them for school.

In the E4Kids study we are directly interested in how participation in early childhood education and care settings contributes to children's learning outcomes and, more generally, prepares children for school and on-going life.

We directly measure children's intellectual abilities using selected tests from the Woodcock-Johnson Tests of Cognition and Achievement.

Participating children have been individually assessed each year for three years in their early childhood and school settings by trained researchers.

The individual tests typically take between 40 minutes and an hour to complete.

In this Bulletin we describe the nature and results of this assessment of children's cognitive functions.

In further work we will be using the data presented here, along with information provided by parents and educators, to assess the contribution of participation in child care and kindergarten to children's learning.

**Collette Tayler and
Karen Thorpe**

Picture from the Early Learning Centre, University of Melbourne.

We would like to provide our heartfelt thanks to those who have assisted E4Kids in undertaking this work:

- Woodcock-Muñoz Foundation
- Kevin McGrew
- Wally Howe
- Psychological Assessments Australia

INSIDE THIS ISSUE

- What is the E4Kids study approach to assessing children's learning?
- What is E4Kids measuring using the Woodcock Johnson III tests?
- How does E4Kids assess children's learning?
- What are the patterns of children's learning in data collected from 2010 to 2012?
- What additional analyses will be undertaken to answer key research questions?

What is the E4Kids study approach to assessing children's learning?

The nature of intelligence

This bulletin is concerned with the measurement of the unobservable (latent) concepts of ability and intelligence. The first measurement of human intelligence, starting with the work of Binet and Simon in the early 1900s, was from the theoretical view that there was an overarching general intelligence that could be measured and expressed as an Intelligence Quotient (IQ) or Mental Age (MA). Over the past century, and most notably in the last 30 years, there has been a rapid shift in our understanding of intelligence. It is now understood that intelligence is in fact made up of multiple aspects that are interrelated. A summary of major developments in measuring intelligence is provided on page 8.

Two broad aspects of intelligence are described as crystallised and fluid. Tests of crystallised intelligence draw on our acquired knowledge and long-term memory, on average peaking in our mid 50s and only then slowly declining. In contrast, tests of fluid intelligence draw on our capacity to solve problems and think logically in new situations, peaking in our mid to late 20s and then declining.

Our theoretical framework

The Cattell-Horn-Carroll (CHC) model of intelligence provides an hierarchical organisation (taxonomy) of the multiple intelligences that make up human ability. This framework integrates the classical models of crystallised and fluid intelligence (Horn-Cattell) with the three-stratum model (Carroll) that orders ability hierarchically: narrow abilities clustered within broad abilities clustered within general intelligence.

The Woodcock-Johnson III (WJ III) is currently the instrument that aligns best with the CHC model and assesses nine broad abilities: crystallized intelligence, fluid intelligence, quantitative reasoning, reading and writing ability, short-term memory, long-term storage and retrieval, visual processing, auditory processing and processing speed. The CHC theory of intelligence is continually growing with the most recent revision including 16 broad abilities, including, for example, kinaesthetic, tactile and olfactory ability and more than 100 narrow abilities within them (McGrew, 2009).

Strengths and limitations

Measuring intelligence is perhaps best described as a work in progress. New ideas are being developed and tested. One of the architects of the CHC theory of intelligence, John Horn, has compared classifying and categorising human abilities and intelligence to 'slicing smoke' (Horn, 1991). Understanding this, WJ III was seen to be the best and most complete set of measures available to the E4Kids study for directly assessing children's abilities and achievements. It is widely used and accepted internationally in early childhood education and care research (Burchinal et al., 2002; Chien et al., 2010; Duncan, 2003; Howell & Kemp, 2010; Howes et al., 2008; Keith & Reynolds, 2010) and is fit for the purpose in the study, though it is of course not the only way of assessing children's abilities and achievements. Other common measures include the WISC III and DAS. These measures show moderate to high correlations with the key clusters of the WJ III (McGrew and Woodcock 2001).

Woodcock-Johnson (WJ III)

WJ III is a set of validated measures of cognition and achievement based on the Cattell-Horn-Carroll (CHC) theory. The Woodcock-Johnson Tests of Cognitive Abilities were first developed in 1977, revised in 1989 and again in 2001 (Mather and Woodcock, 2001a; Mather and Woodcock, 2001b; McGrew, 2009).

The WJ III measure consists of two batteries of tests: Tests of Cognitive Abilities and Tests of Achievement. Some of the tests from the WJ III can be used for children as young as two and normative references (norms) are available from the ages of 2 to 90. Norms provide expected scores on tests for levels of age or education. An Australian adaptation of the test and corresponding norms was produced in 2008.

As mentioned above, the WJ III measures nine broad abilities of the CHC model as well as a range of narrow abilities. It also can be used to calculate clusters that aggregate scores on one or more tests to rate individuals on important conceptual constructs such as academic abilities, for example clustering reading, general intelligence and clinical factors, such as executive processes.

What is E4Kids measuring using the WJ III tests?

E4Kids uses eight tests from the WJ III and these are listed and described in Table 1 below. The percentage of the total number (2,653) of children in the study who completed these tests across the three years of data collection in 2010, 2011 and 2012 are: 89.4%; 84.8%; 87.4%.

Table 1: WJ III tests administered to children in the E4Kids study

Tests used in E4Kids*	Narrow abilities	Description of test	Years in which test administered
Verbal Comprehension (Gc)	Lexical knowledge and language development	Measures four different aspects of language development: picture vocabulary, synonyms, antonyms and verbal analogies	2010, 2011 & 2012
Visual-Auditory Learning (Glr)	Visual auditory associations retrieval	Measures ability to learn, store and retrieve a series of visual-auditory associations	2011 & 2012
Concept Formation (GF)	Fluid reasoning	Measures categorical reasoning based on principles of inductive logic and flexibility of thinking	2010, 2011 & 2012
Visual Matching (Gs)	Processing speed	Measures the speed at which an individual can make visual symbol discriminations	2011 & 2012
Rapid Picture Naming (Gs)	Processing speed	Measures the narrow aspect of naming facility or the speed of direct recall of information from acquired knowledge	2011 & 2012
Understanding Directions (Gc)	Oral language	Measures aspects of comprehension and the more narrow abilities of listening language	2010, 2011 & 2012
Applied Problems (Gq)	Mathematics	Measures capacity to analyse and solve mathematics problems	2010, 2011 & 2012
Sound Awareness (Ga)	Phonological awareness	Measures four aspects of children's phonetic knowledge: rhyming, deletion, substitution and reversal	2011 & 2012

* In brackets are the associated CHC factors as follows: Ga= Auditory processing, Gc= General comprehension, Gf = Fluid reasoning, Glr=Long-term retrieval, Gs= Processing speed, G(q)=Mathematics. The tests administered in E4Kids are only partial indicators of these broader CHC factors.

The basis for the selection of tests from the WJ III was twofold: we applied the criteria that there were a sufficient number of tests to represent the breadth of children's abilities and achievement and, on the other hand, the study met the self-imposed criteria that testing be conducted in less than one hour. This restraint on testing time was for a number of reasons: resource constraints on researcher's time, practical difficulties of managing tests within early childhood settings and schools and limiting the burden on young children's time.

The selection of tests was also constrained by the ages of the children. Whilst the WJ III is used with participants between the ages of 2 and 90, the Tests of Achievement require specific learned skills and abilities, (e.g. writing) which are limited for three and four-year-olds. The selection of tests was therefore also based on specific advice provided in the Examiner's Manual on appropriate tests for preschool age children (Mather & Woodcock, 2001a, 2001b).

In addition, to help make decisions about which WJ III tests to use, we sought and received direct advice from Kevin McGrew (2010), one of the authors on the tests, on how best to cover the broad abilities of children, and their change over time.

The selected tests cover both children's cognitive abilities (Gc, Gf, Glr, Gs, Gq) and cluster to measure general intelligence (BIA) and verbal ability.

Other child outcome data

In further reported measures from this study, we will present data from other methods used to assess children's learning and school achievement, including partnering with government to make use of National Literacy and Numeracy Assessment data (NAPLAN). Parents also provide a number of assessments related to the child's health and wellbeing and the home learning environment that have relevance for their children's learning outcomes. However, this Bulletin focuses on the results of the direct assessments of children's learning undertaken by E4Kids field researchers using selected WJ III tests.

How does E4Kids assess children's learning?

The Verbal Comprehension test is set out below in detail to illustrate how data were collected, while the following page explains how this data was analysed. The four sub-tests of Verbal Comprehension are also the consistent parts of the Verbal Ability cluster. For this Bulletin, we use the term Verbal Ability interchangeably with Verbal Comprehension as Verbal Ability is the higher order concept and more meaningful in relation to a general understanding of children's learning and ability.

Verbal Ability in detail

Verbal Ability measures language development including the comprehension of individual words and the relationships between words. Verbal Ability is measured by the four subtests of the WJ III Cognitive Test of Verbal Comprehension: Picture Vocabulary, Synonyms, Antonyms and Verbal Analogies (described in more detail in the box to the right). Verbal Ability is an indicator of Gc and measures the narrow abilities of Lexical Knowledge and General Verbal Information.

Relevance of the measure

Language and literacy development in children is recognised to be a key to the success of children's longer-term educational and employment outcomes. Low levels of verbal ability, for example, are also strongly associated with low levels of parental literacy/education. Early intervention through support for home-based learning and high quality literacy practices in early childhood education and care settings and schools is accepted as a public policy priority in Australia and internationally. In the early childhood field, the National Quality Standards expected of early childhood education and care settings and the national Early Years Learning Framework and the Victorian Early Years Learning and Development Framework each stress high quality literacy practices, including active investigation, shared thinking and conversations with educators, family members and peers.

In E4Kids, language and literacy practices underpin many aspects of teacher-child interactions, assessed though the Classroom Assessment Scoring System (CLASS), and one specific element on language modelling deals specifically with this.

Test situation

The child sits opposite the researcher (see picture below), is shown one page of material at a time by the researcher and is asked a series of specific questions. The child is asked multiple questions in each of the four sub-tests. Prior to commencing each test, the child is also given two or three sample items to ensure that he or she understands what is being asked. The researcher is asked to record '1' for a correct response and '0' for an incorrect one. There are further detailed instructions to list responses as correct even if there are mispronunciations and differences in verb tense or singular/plural forms, and score incorrect if the child substitutes a different part of speech such as a noun or a verb.

An E4Kids researcher administering the WJ III.

Four subtests of Verbal Abilities

1. Picture Vocabulary

This is a test of lexical (word) knowledge. The child is shown 23 sets of pictures, each with four images, and has to select one of the four images as correct. 'The task requires the person [child] to identify the pictures of familiar and unfamiliar objects….The items become increasingly difficult as the selected pictures appear less frequently in the environment or represent less familiar concepts' (Mather & Woodcock, 2001a, p.11), beginning with a baby as the first test item and ending with a church spire as test item number 23 (Woodcock, McGrew & Mather, 2001, p.7 and p.15).

2. Synonyms

This is a test of vocabulary knowledge (Mather & Woodcock, 2001, p.12). The child is shown a series of 15 words and is asked to tell the researcher another word for each of these. One of the sample items to help the child to understand the task is: 'Tell me another word for 'big'', with acceptable answers listed as 'large, gigantic or huge'. (Woodcock, McGrew & Mather, 2001, p.21)

3. Antonyms

This is a test of a 'counterpart aspect of vocabulary knowledge' (Mather & Woodcock, 2001, p.12). A child is shown a series of 18 words and asked to tell the researcher the opposite of each of them. One of the sample items is: 'Tell me the opposite of 'yes' with the correct response being 'no'. (Woodcock, McGrew & Mather, 2001, p.35).

4. Verbal Analogies

This is a test of the child's 'ability to reason using lexical knowledge' (Mather & Woodcock, 2001, p.12). A child is shown a set of 15 analogies and for each is also given the beginning word of a similar analogy which he or she is asked to complete. In a sample item given to the child, the researcher says: 'finish what I say: a bird flies, a fish…?', with the correct answer being 'swims'. (Woodcock, McGrew & Mather, 2001, p.51)

What are the patterns of children's learning collected from 2010 to 2012?

The data collected on children's Verbal Ability over the three years of the study is summarised in Table 2. The test scores are reported in W units which is a summation of right-wrong answers on the four sub-tests of Verbal Comprehension on a continuous (modified Rasch) scale. The mean value (M) is the average value for all children tested. The standard deviation (SD) measures the spread of scores (variance); the greater the spread, the greater the variation in scores above and below the mean value. In the normal distribution, around 68 per cent of observations fall between one standard deviation above and below the mean.

Table 2: Verbal Comprehension test 2010, 2011 & 2012: E4Kids study children's ages and scores

| Year	Number of children (n)*	Mean (M)	Standard Deviation (SD)	Median	Min	Max
2010 Child age data (months)	2153	48.27	6.95	48	25	73
2010 Verbal Ability data (W)**	2117	450.43	13.76	450	407	495
2011 Child age data (months)	2039	58.63	7.79	58	26	86
2011 Verbal Ability data (W)**	1963	460.05	14.23	460	407	500
2012 Child age data (months)	1953	72.13	7.89	72	38	100
2012 Verbal ability data (W)**	1884	471.70	13.90	472	414	511

* The reported 'n' for child age is that of all children tested. The 'n' reported for Verbal Ability scores is smaller and reflects that some children did not complete all four sub-tests of the Verbal Comprehension test.

**As with other WJ III test scores, 'the W scale for each test is centred on 500 as the approximate average performance of a typical child aged 10 years. The typical range of W abilities within a test is about 430 to 550…'(Jaffe, 2009, p.5). In the normative sample, the W scores for Verbal Ability for ages 2 to 17 ranges between 426 and 526.3 (McGrew, personal communication, 2013).

In the E4Kids Study the mean of Verbal Ability across the three years of data collection range between a low of 450.53 in 2010 and a high of 471.70 in 2012.

The underlying data are presented in Figure 1. The coloured dots represent individual test scores, with each colour indicating the three years of data collection: 2010, 2011 and 2012. The three coloured lines represent a linear (OLS) trend for each of the three years of data collection. The black dots represent Australian normative values for the same test to age 8 years. As indicated by the distribution of the individual scores on this test, the youngest children tested in E4Kids were just over two years of age and the oldest between seven and eight years of age. The shaded areas outline the data falling within one standard deviation of the norm values.

Figure 1: Verbal Comprehension/Ability Scores: E4Kids Study and Normative Values
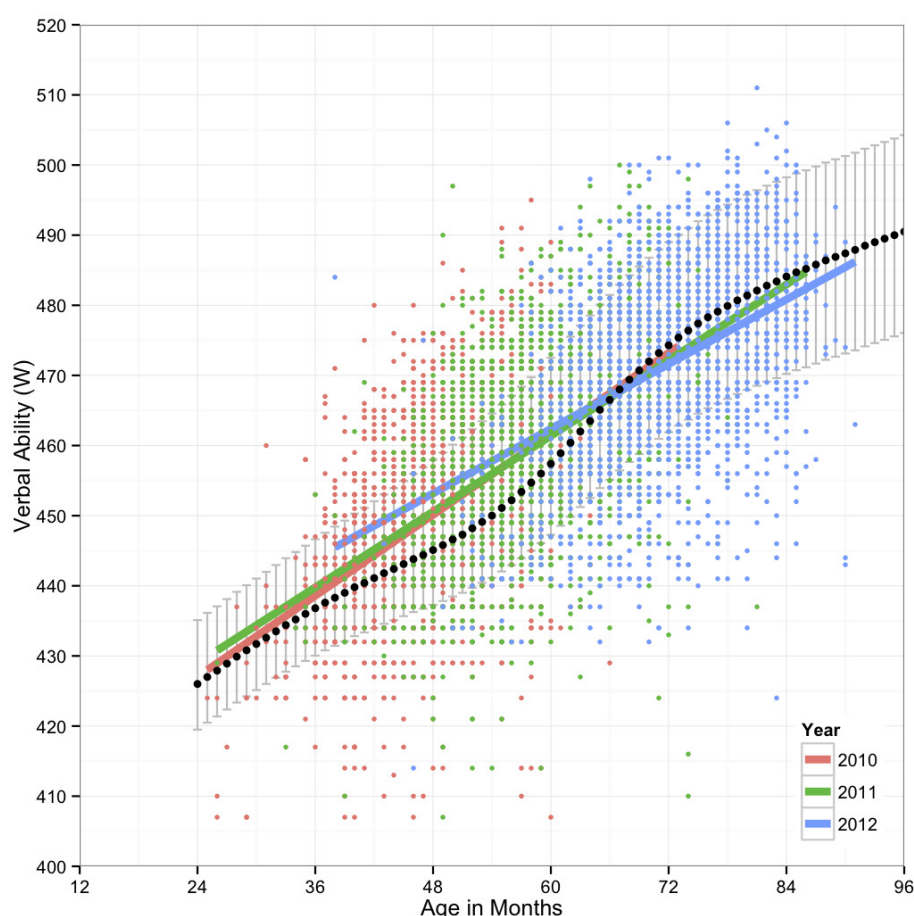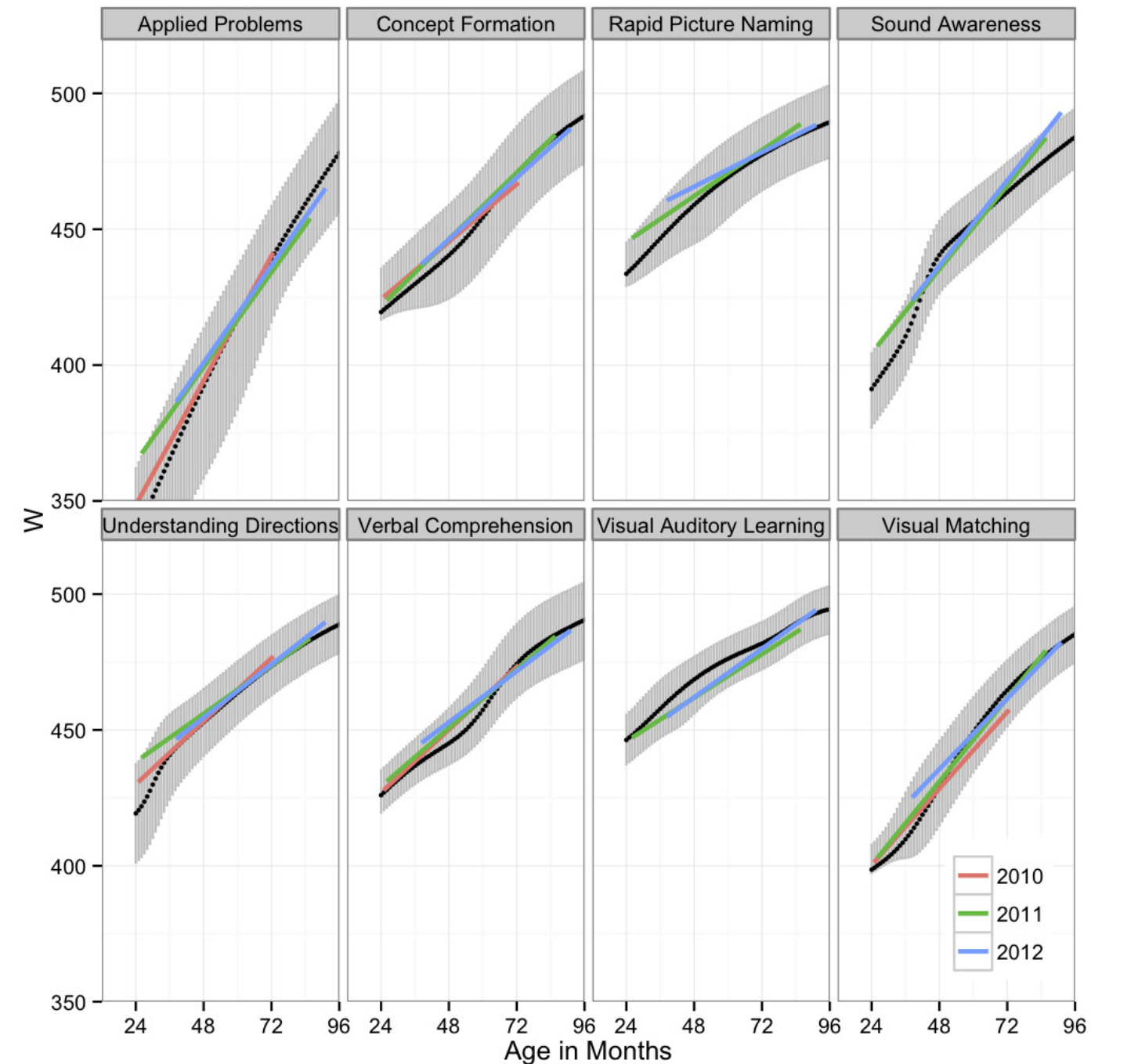

Technical note: The Australian normative values are provided by McGrew (personal communication, 2013) and include a small amount of US normative data to supplement the Australian data.

What are the patterns of children's learning collected from 2010 to 2012? (cont.)

The same form of illustrating E4Kids data relative to normative data that was used for Verbal Ability is provided across the full eight test scores, including Verbal Ability (see Figure 2 below). Please note that for four of these figures there are only two rounds of data collection, 2011 and 2012, as indicated in Table 1 (page 3).

Figure 2: WJ III test scores in E4Kids and normative scores on all tests used

Technical note: Figure 2 presents E4Kids and normative Australian data from age 2 years to age 8 years.

Normative Australian data

The Australian normative data for the eight tests to age 17 are provided in Figure 3.

Interestingly, the mean values of the test scores increased rapidly in about the first eight years of life, intersect when children are between about 8 and 10 years, and then diverge into quite different growth patterns. This is part of a larger trend where growth in fluid intelligence measures peak in the 25 to 30 age group and measures of crystallised intelligence peak in the mid 50s.

Figure 3: Ten WJ III Australian normative test data scores to age 17 years
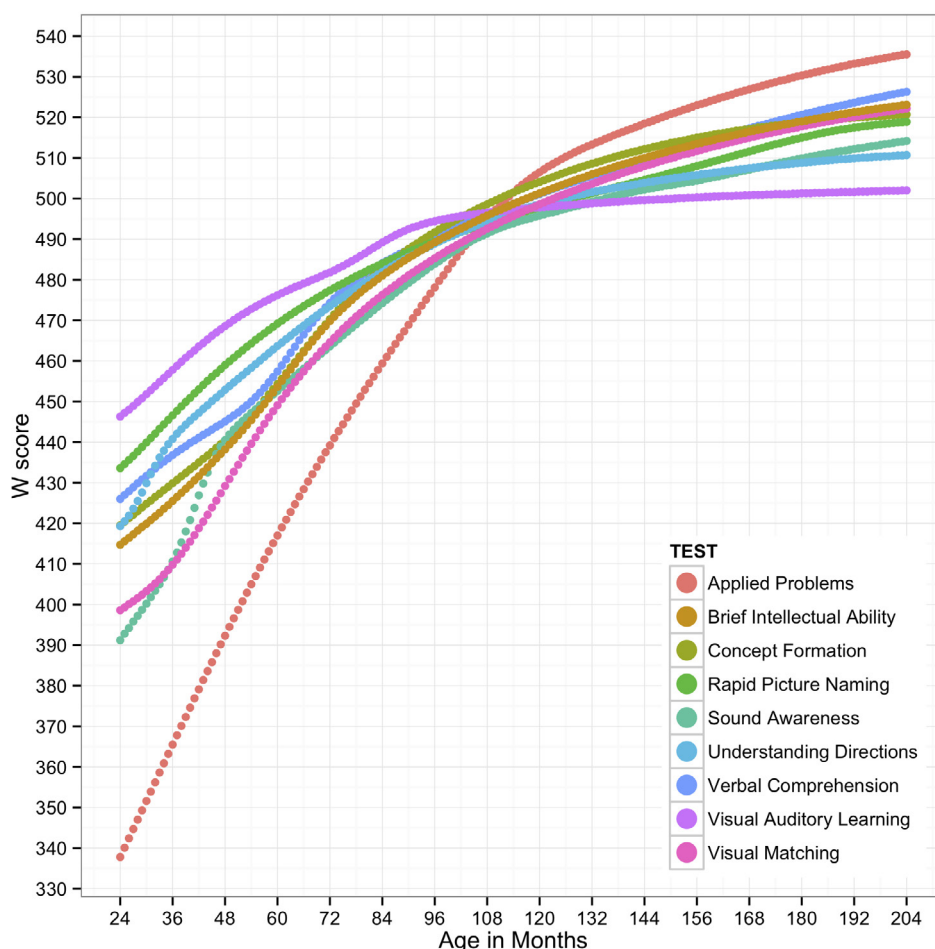
What additional analyses will be undertaken to answer key research questions?

In answering study research questions, more finely attuned analysis will focus on aspects of children's participation and service quality:

- What are the impacts of 'dosage' of Early Childhood Education and Care (ECEC) programs on child learning outcomes?

- What are the impacts of program quality on child learning outcomes, as measured by ECERS-R (Early Childhood Environment Rating Scale-Revised Edition) and CLASS measures?

In assessing the impact of ECEC programs on child outcomes, it will be necessary to establish the relative influence of other factors known to influence child outcomes (and of interest in their own right), including:

- What are the impacts of a range of family circumstances, such as parental Socio Economic Status (SES) and family stresses?

- What are the impacts community factors such as community SES and parental ratings of neighbourhoods on child outcomes?

- What are the impacts of factors more intrinsic to the child, such as temperament or having a major health problem or disability?

Key historical developments in our understanding of intelligence are summarised by Berk (2013) and include:

- The development of an intelligence quotient of mental ability over age (Stanford-Binet Intelligence Scale): originally initiated for the purpose of identifying children of lower ability in schools and excluding them from general classes on the basis that they were holding back more capable students.

- Early developments in factor analysis, that put forward a general underlying intelligence, made up of a cluster of abilities as well as unrelated factors.

- Later developments in factor analysis that identified two types of intelligence: crystallised intelligence which draws on one's understanding and competence in one's own culture, and fluid intelligence which depends more on general information processing skills, including speed in analysing information and working memory capacity. Whilst crystallised intelligence shows improvement through adulthood and declines only in older age, fluid intelligence has shown to peak in the mid-twenties and then decline.

- Further developments in factor analysis in where a three tiered system was envisaged comprising a general intelligence made up of eight broad factors, including fluid and crystallised intelligence, and eight specific factors including, sequential reasoning, vocabulary knowledge, creativity and reaction time (see Carroll's three-stratum theory of intelligence described in Berk (2013, p. 322))

- A more recent development is Sternberg's notion of three broad interacting intelligences: (1) analytic intelligence or information processing skills, (2) creative intelligence, the capacity to solve novel problems and (3) practical intelligence, the application of intellectual skills in everyday situations (Berk 2013, p. 323-324)

- Gardner's theory of intelligences further extends the notion of information processing skills into eight separate domains: linguistic, logico-mathematical, musical, spatial, body-kinaesthetic, naturalist, interpersonal and intrapersonal (Berk 2013). Whilst the biological bases for these different types of intelligence have not been clearly demonstrated, the experience of the Master of Teaching (Early Childhood) course at the University of Melbourne is that they have proved very helpful in planning lessons in early childhood settings to develop a full range of abilities in children.

Picture from the Early Learning Centre, University of Melbourne.

Conclusion

This Bulletin provides an overview of the main way in which this study measures children's intellectual and learning outcomes using the WJ III. It also reports the main patterns of scores over three years of data collection in E4Kids. The data are presented in the context of Australian normative scores over the first seventeen years of the same eight tests.

In E4Kids, WJ III provides robust data on child intellectual and learning outcomes. The data are critical to answering the key research question of the study: what are the impacts of participation in ECEC programs on children's learning and development? WJ III is used internationally, recognised to be robust, and will make E4Kids result comparable with other major studies.

We will also be drawing on parental reports of children's social and behavioural development and school achievement data, in particular NAPLAN data, for 2014 and 2015.

References

Berk, L. (2013). *Child Development* (9th ed.). New Jersey, US: Pearson Publishing.

Burchinal, M., Peisner-Feinberg, E., Pianta, R., & Howes, C. (2002). Development of Academic Skills from Preschool Through Second Grade: Family and Classroom Predictors of Developmental Trajectories. *Journal of School Psychology*, 40(5), 415-436.

Chien, N., Howes, C., Pianta, R., Burchinal, M., Ritchie, S., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2010). Childrens' classroom engagement and school readiness gains in prekindergarten. *Child Development*, 81(5), 1534-1549.

Duncan, G. (2003). Modeling the Impacts of Child Care Quality on Children's Preschool Cognitive Development. *Child Development*, 74(5), 1454-1475. doi: 10.1111/1467-8624.00617

Horn, J. (1991). Measurement of intellectual capabilities: A review of theory. *Woodcock-Johnson technical manual*, 197-232.

Howell, S., & Kemp, C. (2010). Assessing preschool number sense: skills demonstrated by children prior to school entry. *Educational Psychology*, 30(4), 411-429.

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-Kindergarten programs. *Early Childhood Research Quarterly*, 23(1), 27-50.

Jaffe, L. (2009). Development, interpretation and application of the W score and the Relative Proficiency Index, in *Woodcock Johnson III Assessment Service Bulletin, Number 11*. Rolling Meadows, US: Riverside Publishing.

Keith, Z., & Reynolds, M. (2010). Cattell-Horn-Carroll Abilities and Cognitive Tests: What We've Learned from 20 Years of Research. *Psychology in the Schools*, 47(7), 635-650.

Mather, N., & Woodcock, R. (2001a). *Woodcock Johnson III, Test of cognitive abilities, Examiner's manual*, Standard and extended batteries. Itasca, US: Riverside Publishing.

Mather, N., & Woodcock, R. (2001b). *Woodcock Johnson III, Test of achievement, Examiner's manual*, Standard and extended batteries. Itasca, US: Riverside Publishing.

McGrew, K. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37, 1–10.

McGrew, K. (2010). Reflections on 35 Years of Intelligence Testing. *E4Kids Seminar*, 09/2010. http://web.education.unimelb.edu.au/E4Kids/news/pdfs/20101004_aus34.pdf

McGrew, K., & Woodcock, R. (2001). *Woodcock-Johnson III, technical manual*. Itasca, US: Riverside Publishing.

Woodcock, R., McGrew, K. & Mather, N. (2001). *Woodcock-Johnson III, Tests of Achievement, Standard Test Book Tests 1-12*, Form B. Itasca, US: Riverside Publishing.

To subscribe to E4Kids mailing list for future issues, please visit our webpage at www.e4kids.org.au/mailing_list

Research Collaboration

MELBOURNE GRADUATE SCHOOL OF EDUCATION
THE UNIVERSITY OF MELBOURNE

QUT
Queensland University of Technology

Victoria State Government
Department of Education and Early Childhood Development

Queensland Government

Australian Government
Australian Research Council