



UvA-DARE (Digital Academic Repository)

Validity and Reliability of the Early Development Instrument in Indonesia

Brinkman, S.A.; Kinnell, A.; Maika, A.; Hasan, A.; Jung, H.; Pradhan, M.

DOI

[10.1007/s12187-016-9372-4](https://doi.org/10.1007/s12187-016-9372-4)

Publication date

2017

Document Version

Final published version

Published in

Child Indicators Research

License

Article 25fa Dutch Copyright Act (<https://www.openaccess.nl/en/in-the-netherlands/you-share-we-take-care>)

[Link to publication](#)

Citation for published version (APA):

Brinkman, S. A., Kinnell, A., Maika, A., Hasan, A., Jung, H., & Pradhan, M. (2017). Validity and Reliability of the Early Development Instrument in Indonesia. *Child Indicators Research*, 10(2), 331-352. <https://doi.org/10.1007/s12187-016-9372-4>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Validity and Reliability of the Early Development Instrument in Indonesia

Sally A. Brinkman^{1,2} · Angela Kinnell^{1,2} ·
Amelia Maika^{2,3} · Amer Hasan⁴ · Haeil Jung⁵ ·
Menno Pradhan^{6,7}

Accepted: 24 January 2016 / Published online: 6 February 2016
© Springer Science+Business Media Dordrecht 2016

Abstract There is increasing interest from international organizations and the research community to use internationally comparable instruments that in turn foster global understanding while providing evidence for local and international policy development. In the field of early childhood, international comparisons have traditionally been limited to indicators such as infant or child mortality and anthropometric data such as stunting and wasting. However, there has been gradual interest in developing international measures that can be used to compare and monitor the holistic development of children. Using both the short and standard versions of the Early Development

✉ Sally A. Brinkman
sallyb@icmr.uwa.edu.au

¹ Telethon Kids Institute, The University of Western Australia, PO Box 855, West Perth 6872, Western Australia

² School of Population Health, The University of Adelaide, Level 7, Terrace Towers, 178 North Terrace, Adelaide 5005, South Australia

³ Department of Sociology, Gadjah Mada University, SosioYustisia Bulaksumur, Yogyakarta 55281, Indonesia

⁴ Education Global Practice, The World Bank, 1818 H St NW, Washington, DC 20433, USA

⁵ Department of Public Administration, Korea University, 145 Anam-Ro, Seongbuk-Gu, Seoul 02841, Korea

⁶ Amsterdam Institute for International Development, Faculty of Economics and Business Administration, VU University Amsterdam, de Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

⁷ Faculty of Economics and Business Administration, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

Instrument (EDI), this paper reports on the process of adaptation of the EDI in Indonesia. Further, it explores the content and construct validity, internal consistency, inter-rater reliability and predictive validity of the EDI using a number of measures including the Strengths and Difficulties Questionnaire, the Dimensional Change Card Sort, and school-based tests of language, mathematics and cognitive performance, collected from a number of informants (caregivers, teachers, and children). We report on data for two cohorts of children: the “younger cohort” were approximately 1 year old ($N=3116$) and the “older cohort” were approximately 4 years old ($N=3251$) at Time 1. Both cohorts were followed up approximately 4 years later, at Time 2. This study finds that the EDI shows moderate validity and reliability in poor communities in Indonesia and highlights some of the difficulties associated with adapting western instruments for non-western cultures and contexts.

Keywords Child Development · Early Development Instrument (EDI) · Validity · Reliability · Indonesia

1 Introduction

A country’s progress is usually assessed by collecting and analysing various basic health and education statistics, including infant, child and maternal mortality rates, breastfeeding rates, immunisation rates and primary school enrolments and attendance. While there is no doubt about the fundamental importance of these statistics, there is an increasing awareness that additional information is required to assess not just basic health status, but how well children are developing. The measurement of child development is also important considering the implications for monitoring that are highlighted by the United Nations Convention on the Rights of the Child (Bernard van Leer Foundation 2006; UNICEF 2009). Each of the countries which are signatories to the Convention are responsible for providing children with the opportunities necessary to develop physical, cognitive, social and emotional capacities in early life.

Measuring child development and comparing it between countries has its complexities. Child development is influenced by culture, language and theory, and thus the concept can vary across place, culture, language and research tradition, making international comparability difficult (Hambleton et al. 2005). The definitions of birth and death are very consistent across countries but assessing an indicator like breastfeeding requires strict adherence to internationally agreed upon definitions (e.g., at what point does exclusive breastfeeding become complementary). More complex again are indicators such as psychological assessment, educational assessment, and child development and agreeing on definitions of these indicators requires significant consideration prior to making any cross country or culture comparisons.

The International Test Commission (ITC) was formed in response to the methodological challenges of adapting and validating test instruments (usually psychological and educational) for comparisons between countries. The Guidelines (International Test Commission (ITC) 2000) stipulate the various steps that should be taken before instruments can be used to reliably compare results across countries. The ITC Guidelines can be equally applied to the field of child development and have been used by the developers of the Early Development Instrument (EDI) (Janus and Offord

2007), a measure of child development, to ensure it is appropriate for use in a range of different countries.

1.1 The Early Development Instrument

The EDI is a measure of child development which, to date, has been used in more than 20 countries, ranging from small scale studies through to regional and/or population coverage (e.g., in Australia, Canada, United States, Mexico and Chile) (Brinkman 2009). The EDI covers five key domains of child development: Physical health and well-being, Social competence, Emotional maturity, Language and cognitive skills, and Communication skills and general knowledge. In most countries the EDI is completed by teachers in the child's first formal/full-time year of school. While the EDI is completed for individual children in a class, the scores are aggregated to school, neighbourhood, regional or country level. The interpretation occurs at group level, based on the analyses of all children and with reference to the distribution of all scores. The EDI's intended use is as a population measure to compare groups of children rather than as an individual diagnostic tool.

The EDI's guidelines ("How to use the EDI," 2015) outline the steps to be undertaken before the EDI is used in a new country and prior to claiming that the instrument is valid for use within that setting. As noted above, these guidelines are consistent with the ITC Guidelines (Hambleton et al. 2005) and adhere to a universalist approach (Herdman et al. 1998), making no a priori assumptions about comparability. In accordance with these various guidelines, the EDI needs to be validated and tested for reliability within a country prior to being compared across countries. This paper is the first to report on the reliability and validity of the EDI in Indonesia. There is ongoing validation of the EDI in many countries around the world and results to date have shown the EDI to have good validity and reliability.

Concurrent validity in other countries has been previously demonstrated via correlations between the domains of the EDI and other related measures of child development. For example, the Social competence and Emotional maturity domains of the EDI have been found to correlate moderately to strongly with the domains of the Strengths and Difficulties Questionnaire (SDQ, see section 2.1.2) in Canada (Janus et al. 2007). Notably, in Canada, Australia and Hong Kong moderate correlations have also been identified between the Language and cognitive skills domain of the EDI and various other measures of cognitive abilities including the Peabody Picture Vocabulary Questionnaire (PPVT), the Who am I? test, and the Hong Kong Early Child Development Scale (HKECDS) (Ip et al. 2013; Janus et al. 2011; Janus and Offord 2007).

The predictive validity of the EDI has also been demonstrated. In Canada, the Language and cognitive domain, the Communication skills and general knowledge and the Physical health and wellbeing domains showed to be predictors of later school achievement (Forget-Dubois et al. 2007). Another Canadian study found the Physical health and well-being domain to be a good predictor of Grade 2 visual-motor integration ability, the Social competence and Emotional maturity domains to predict Grade 2 SDQ scores, and the Language and cognitive skills and Communication and general knowledge domains to predict Grade 2 scores on the Detroit Test of Learning Aptitude (DTLA) (Janus et al. 2007). Australian data has shown that a child's score on the EDI

domains at 4 years old is a significant predictor of their later (Grades 3, 5, and 7) school assessed literacy and numeracy performance (Brinkman et al. 2013).

The inter-rater reliability for the EDI has been looked at in several studies. Generally, the inter-rater reliability between two different teachers or professionals (e.g., kindergarten teacher and early childhood educator) was greater than the similarity between ratings of the child's parent and their teacher. This finding is consistent with the finding that teachers are a better judge of children's abilities than are their parents (Sommer et al. 2008). Correlations between the EDI scores of two different teachers have been found to range between 0.53 and 0.8 while the correlations between parent and teacher ratings have been found to range from 0.36 to 0.64 (Janus and Offord 2007).

The internal consistency of the domains of the EDI has been examined using data from a range of locations and has shown remarkably similar results - Canada (range 0.76–0.96) (Forer and Zumbo 2011; Forget-Dubois et al. 2007; Janus et al. 2007), Australia (0.80–0.95), USA (0.82–0.96), Jamaica (0.64–0.95) (Janus et al. 2011), Scotland (0.78 - <0.9) (Woolfson et al. 2013) and China (0.70–0.95) (Ip et al. 2013). In each of these studies, internal consistency was lowest for the Physical health and well-being domain in comparison to the other four domains.

1.2 Indonesian Context

Indonesia is classified as a lower middle income country and continues to show strong economic growth (World Bank 2012). However, living conditions in Indonesia have not improved commensurate to this growth and just under 12 % of the population lives below the poverty line (World Bank 2012, 2013). In a population of 247 million (2012), this means that over 28 million people are living below the poverty line in Indonesia with a further 74 million people (30 %) considered vulnerable to poverty (World Bank 2013). The majority of people living below the poverty line reside in rural areas of Indonesia and there is large inequality between the rich and the poor (World Bank 2013).

In recent years, there has been an increase in the use of maternal child health services in Indonesia with maternal and child mortality rates decreasing. The proportion of births attended by a professional health provider has increased to nearly 75 % (Mize et al. 2010). The maternal mortality rate dropped from 340 to 220 maternal deaths per 100,000 live births and the infant mortality rate dropped from 38 to 27 deaths per 1000 live births between 2000 and 2010. However, coupled with the inconsistent provision of health services and financial constraints on people's ability to access services, Indonesia's maternal and infant mortality rates remain the highest in East Asia (Mize et al. 2010; World Bank 2010). There is also wide variation in educational outcomes in Indonesia. While data indicate that all Indonesian children start primary school, poorer children and those living in rural areas complete fewer years of schooling and have higher drop out and repetition rates than those who are less disadvantaged (Central Board of Statistics of Indonesia 2010; Hasan et al. 2013).

In response to poor educational outcomes the Ministry of National Education (MoNE) gave its support to a community-based early childhood education and development project which aimed to reach 738,000 children in 50 districts over 5 years. The project provided block grants to communities which tended to be spent on establishing centre-based playgroups. In addition, the program funded training of community-based

teachers in promoting child development. The financial support for this program was provided through a country loan from the World Bank with additional funding from the Dutch Government. More detail regarding this project can be found online (<http://www.worldbank.org/projects/P089479/early-childhood-education-development-project?lang=en>).

In an effort to understand whether this program improved children's development and readiness for primary school, the MoNE undertook an impact evaluation, generally known as the Indonesian Early Childhood Education and Development (ECED) Impact Evaluation, with technical support from the authors and agencies listed on this paper. This evaluation used a randomized and matched control design that allowed for comparison between similar communities that received the project at different points in time (ISRCTN76061874. For full details on this evaluation, please consult the study protocol - Pradhan et al. 2013). When designing the Impact Evaluation there was a commitment to use measures that could be compared internationally. Prior to this study there were no known instruments (locally or internationally developed) that had been validated to measure child development and readiness for primary school in Indonesia

2 Materials and Methods

2.1 Participants

The data reported in this paper were part of the larger Indonesia ECED Impact Evaluation. We report on data for two cohorts of children. The "younger cohort" were approximately 1 year old ($N=3116$) and the "older cohort" were approximately 4 years old ($N=3251$) at Time 1. Both of these cohorts were followed up approximately 4 years later, Time 2, when they were approximately 5 years and 8 years old, respectively.

As noted above, the work reported in this paper is a component of a larger Impact Evaluation. At Time 1, the key measure of interest in this paper, the Early Development Instrument, was only completed for the older cohort as the younger cohort were at that time not old enough to be the subject of the EDI. Thus for this paper, no data are reported for the younger cohort at Time 1. The EDI was completed for both cohorts of children at Time 2.

Participants were all residents in villages in rural Indonesia. A measure of household wealth was calculated based on information obtained from household interviews. This information included the assets possessed by the household and the materials used in the construction of the homes in which they lived. Principal components analysis was used to construct an index of household wealth which was then used in the analyses. Households were generally poor with over 30 % of families having received cash transfers and over 70 % having received food assistance before Time 1. In terms of housing conditions; 92 % of households had electricity, 30 % had drinking water available inside their house, 50 % were linked into sewerage infrastructure and 43 % had their own toilet.

At Time 1, more than 50 % of the primary caregivers sampled had not completed junior high school and approximately 15 % were unable to read a literacy check sentence ("Parents must give attention to their children" / "*Orang tua harus memberikan perhatian kepada anaknya*"). Anthropometrics were also measured and

showed that 36 % of sample children were stunted and 30 % wasting. Table 1 provides detail on the key socio demographic characteristics of the sample.

2.2 Instruments

2.2.1 The Early Development Instrument (EDI)

The Early Development Instrument (EDI) consists of 104 items, covering five domains of child development. The EDI is also available in short form; a 47 item instrument. As the EDI was being used in a suite of instruments there was concern from the outset that the questionnaires may become too time-consuming for respondents. It was therefore decided to use the short form of the EDI at Time 1. It is acknowledged that the author of the EDI generally does not recommend the use of the short form EDI prior to analyses of all the items by piloting the full EDI in a country first (Janus and Duku 2005). At Time 2, the size of the suite of questionnaires was significantly reduced which allowed for the use of the long version of the EDI for the younger cohort.

2.2.2 The Strengths and Difficulties Questionnaire (SDQ)

The Strengths and Difficulties Questionnaire (SDQ) (Goodman 2001) is an informant-based assessment of a child's behaviour. The SDQ is comprised of 25 items across five scales; Emotional symptoms, Conduct problems, Hyperactivity/

Table 1 Key socio demographic characteristics of the sample

	Time 1 - older cohort		Time 2 – older cohort		Time 2 – younger cohort	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Socio-demographic characteristics						
Child gender						
Male	1615	49.68	1508	50.02	1427	51.20
Female	1636	50.32	1507	49.98	1360	48.80
Stunting						
Not stunted	2015	63.91	2194	73.55	2006	73.56
Moderately stunted	822	26.07	653	21.89	579	21.23
Severely stunted	316	10.02	136	4.56	142	5.21
Wasting						
Not wasting	2254	70.13	1852	62.53	1823	66.48
Moderately wasting	743	23.12	794	26.81	676	24.65
Severely wasting	217	6.75	316	10.67	243	8.86
Primary caregiver highest education level						
Elementary school or below	1569	52.47	1413	51.80	1427	48.11
Junior high school	746	24.95	685	25.11	764	25.76
Senior high school	510	17.06	480	17.60	590	19.89
Tertiary education	165	5.52	150	5.50	185	6.24

inattention, Peer problems and Prosocial behaviour. The SDQ Total difficulties score may also be calculated by summing a child's scores on the first four 'problem' scales. A high score on these four scales and the SDQ total difficulties score is a negative outcome, indicating that the child has been assessed as having more behavioural problems than a child with a lower score. The Prosocial behaviour scale is usually scored such that high scores indicate better development, however, for consistency with the presentation of results for the other domains, results for the Prosocial behaviour scale in this paper have been reverse scored such that high scores indicate poorer Prosocial behaviour.

The SDQ is widely used around the world and had already been translated for Indonesia (Goodman 2005) prior to this project. There is a considerable amount of validation data for the SDQ in various other countries, much of which is available via the official SDQ website (<http://www.sdqinfo.com>). These data show consistent factor structure and psychometric properties satisfactory for the use of the SDQ in those countries, for example the SDQ has been shown to have satisfactory reliability with mean internal consistency of Cronbach's alpha of 0.71 across scales for parents (Goodman 2001; Goodman et al. 2000; Hawes and Dadds 2004; Hayes 2007). However, no validation data on the use of the SDQ in Indonesia has been published.

Given that one purpose of the SDQ is to screen children for emotional and behavioural problems, the SDQ is most likely to be related to the Social competence and Emotional maturity domains of the EDI. Thus the SDQ provides the opportunity to test the concurrent construct validity of these EDI domains. The SDQ was completed by caregivers of the older cohort at Time 1 and Time 2 and by the caregivers of the younger cohort at Time 2.

2.2.3 The Dimensional Change Card Sort (DCCS) Task

The Dimensional Change Card Sort task (Frye et al. 1995; Zelazo 2006) is a test of executive function completed by the children themselves. In this task, a child is asked to sort cards which vary in both colour (red and blue) and shape (traditionally, a rabbit and a boat but changed to a cat and motorbike for Indonesian children because of their greater familiarity with these objects). Children are first asked to sort the cards according to one dimension (e.g., colour) before switching to sorting them by the other dimension (e.g., shape). International research has shown that most 3 year olds can sort the cards according to one dimension but only about 19 % of 3 year olds can switch to sorting by the other dimension, even when they are instructed to do so before each trial. Most 4 year olds can successfully make the dimensional shift (Carlson 2005; Zelazo 2006). The advanced version of the task includes borders on some cards. A different sorting rule applies depending on whether or not a border is present; if the card has a border the child must sort by one dimension (e.g., colour) and must sort by the other dimension (e.g., shape) if a border is not present. Most 4 year olds and about 50 % of 5 year olds struggle with the border version of the task (Carlson 2005; Zelazo 2006). Similar age trends have been observed with the DCCS task with preschoolers in China and the U.S. (Sabbagh et al. 2006). The DCCS task was completed by the older cohort at Time 1 and by the younger cohort at Time 2.

2.2.4 Child School Tests – Language, Mathematics, and Cognitive Performance

School-based tests were developed specifically for this project based on learning standards in Indonesian schools. Two tests were developed, one for completion by 6 and 7 year old children (the age of the typical Indonesian child at school entry) and the other for completion by 8 and 9 year old children. Children completed the test on the basis of their age, rather than their school grade, as the latter was a potentially endogenous variable in the Impact Evaluation in that children exposed to the project may have started school later than others of the same age. Children completed the tests in a classroom under the guidance of a member of the data collection team; class teachers were not present. Questions were read aloud and any queries children had about the questions were answered.

Each test was divided into three broad sections – language (Bahasa Indonesia), mathematics, and a test of cognitive ability modelled on the Raven's Progressive Matrices. Language questions (23 questions for 6–7 year olds and 21 questions for 8–9 year olds) included questions about association (matching a picture's colour with the written name of the colour – 6–7 year old test only), selecting the picture whose name began with a different letter to other pictures, writing the name of everyday items (e.g., hat, window and house), and comprehension of written text. The mathematics section (15 questions for 6–7 year olds and 25 questions for 8–9 year olds) included some multiple choice questions involving simple addition and subtraction (using differing numbers of pictures, rather than numerals), descriptive mathematics problems, and ordering sequences of numbers from largest to smallest and vice versa. Raven's Progressive Matrices are a test of nonverbal cognitive abilities. Children are presented with an image that is missing a small section and are asked to select the missing piece from six options, based on colour, texture, and orientation, among other factors. Younger children were given one Raven's-like problem per page (a total of 14 questions) while the older children were given two Raven's-like problems per page (a total of 18 questions).

In total, 52 questions were asked of the younger children and 64 questions were asked of the older children. There was some overlap in questions between the two tests with the test for older children omitting the easier questions at the beginning of each section of the 6 and 7 year old test. The school-based tests were completed by the older cohort at Time 2.

2.3 Procedure

All data reported were collected as part of the Indonesian ECED Impact Evaluation. The methodology for this Evaluation will be briefly summarised here and is explained in greater detail elsewhere (Alatas et al. 2013; Hasan et al. 2013; Pradhan et al. 2013). The data analysed in this paper come from two rounds of data collection – Baseline in 2009 (Time 1) and Endline in 2013 (Time 2). Data were collected across nine districts of Indonesia which were selected by MoNE and the World Bank due to their low service availability, low participation in ECED services, a demonstrated commitment to upgrading ECED services, and high poverty rates. These districts were: Sarolangun, North Bengkulu, East Lampung, Majalengka, Kulon Progo, Rembang, Ketapang, Sidenreng Rappang (Sidrap), and Central Lombok.

A range of data was collected from a number of informants. Of relevance to this paper is that, at Time 1, caregivers¹ of the older cohort completed the short form EDI and the SDQ for their children and the children themselves completed the Dimensional Change Card Sort (DCCS; see section 2.1.3) Task. The Time 1 data for this cohort are used to determine internal consistency and concurrent construct validity of the short-form EDI.

At Time 2 all children in the older cohort completed school-based tests covering language, mathematics and cognitive abilities (these measures are explained in greater detail in section 2.1.4). These data are used to determine the predictive validity of the short form EDI (i.e., how well does the EDI collected at Time 1 predict school assessments at Time 2?).

An additional advantage for the assessment of reliability of the EDI was that for the younger cohort at Time 2 both their caregiver and their teacher (if they were enrolled in an ECED service) completed the full EDI (i.e., not the short-form). These data provide for the assessment of inter-rater reliability. As the SDQ and the DCCS were also completed at Time 2 for the younger cohort, this provides another opportunity to test the concurrent construct validity of the EDI, however in this circumstance the validity of the long form EDI.

3 Results

3.1 Descriptive Statistics

Tables 2 and 3 present the descriptive statistics for the instruments used in this study. The results are as expected with the EDI domains showing a skew to the higher (better) scores which is consistent with results found in other countries. The scores on the SDQ, if anything, are a little higher (poorer results) than would be expected compared to international norms, however the patterns across the scales are consistent.

3.2 Content Validity of the EDI and its Adaptation for Indonesia

The content validity of the short form of the EDI in Indonesia was established over a number of years (from 2006 to 2009) through meetings, focus groups and several pilots in the field. First, staff from the Government of Indonesia with expertise in early child development attended meetings with the research team and translators. Feedback from these meetings resulted in a set of questions related to the child's knowledge of Muslim prayers and religious practices being included in the questionnaires. No other modifications were made at this stage. The instrument was then piloted in several poor urban and rural villages near Majalengka (West Java) in March 2007. Eight completed questionnaires were returned.

The pilot revealed that respondents preferred simple yes or no response options rather than the usual 5 point Likert scale response options (i.e., "sometimes agree"

¹ The EDI is usually completed by a child's class teacher. However, in Indonesia a child generally does not begin school until they are 7 years of age meaning this cohort had not yet begun school. Thus, the EDI was completed for this cohort by their caregiver as part of the overall suite of questionnaires.

Table 2 Descriptive statistics for child development indicators

Instrument/Variable	N	Minimum	Maximum	Mean (SD)	Number of items
Child Indicators					
Time 1 (older cohort)					
EDI Physical health & well-being	3251	1.67	10.00	8.57 (1.33)	6
EDI Social competence	3251	0	10.00	7.82 (1.38)	12
EDI Emotional maturity	3251	0	10.00	6.57 (1.53)	12
EDI Language & cognitive development	3248	0	10.00	3.14 (2.04)	12
EDI Communication & general knowledge	3251	0	10.00	9.58 (1.24)	3
SDQ Emotional symptoms	3251	0	10.00	3.70 (2.04)	5 ^a
SDQ Conduct problems	3251	0	10.00	3.48 (1.90)	5
SDQ Hyperactivity/inattention	3251	0	10.00	5.04 (1.31)	5
SDQ Peer problems	3250	0	9.00	2.60 (1.52)	5
SDQ Pro-social behaviour ^b	3241	0	10.00	3.61 (1.90)	5
SDQ Total Score	3249	2	33.00	14.81 (4.52)	20
Time 2 (younger cohort)					
EDI (Caregiver) Physical health & well-being	2771	0	10.00	7.83 (1.44)	11
EDI (Caregiver) Social competence	2774	0	10.00	6.53 (1.51)	26
EDI (Caregiver) Emotional maturity	2777	1.17	9.83	5.90 (1.28)	30
EDI (Caregiver) Language & cognitive development	2783	0	10.00	4.68 (2.61)	26
EDI (Caregiver) Communication & general knowledge	2784	0	10.00	6.63 (2.11)	8
EDI (Teacher) Physical health & well-being	1255	2.27	10.00	8.22 (1.28)	11
EDI (Teacher) Social competence	1254	0.8	10.00	6.94 (1.60)	26
EDI (Teacher) Emotional maturity	1253	1.17	10.00	6.41 (1.36)	30
EDI (Teacher) Language & cognitive development	1254	0	10.00	6.98 (2.38)	26
EDI (Teacher) Communication & general knowledge	1255	0	10.00	6.17 (1.90)	8
SDQ Emotional symptoms	2791	0	10.00	3.97 (2.16)	5
SDQ Conduct problems	2791	0	10.00	3.87 (2.03)	5
SDQ Hyperactivity/inattention	2791	0	10.00	4.85 (1.43)	5
SDQ Peer problems	2790	0	9.00	2.43 (1.52)	5
SDQ Pro-social behaviour ^a	2789	0	10.00	3.13 (1.87)	5
SDQ Total Score	2790	2.00	32.00	15.13 (4.78)	20
Time 2 (older cohort)					
Short EDI Physical health & well-being	3247	2.50	10.00	9.11 (1.18)	6
Short EDI Social competence	3247	0	10.00	7.72 (1.40)	12
Short EDI Emotional maturity	3247	1.25	10.00	6.83 (1.42)	12
Short EDI Language & cognitive development	3246	0	10.00	9.48 (1.39)	12
Short EDI Communication & general knowledge	3247	0	10.00	7.95 (2.19)	3
Long EDI Physical health & well-being	3247	2.00	10.00	8.76 (1.22)	11
Long EDI Social competence	3247	0	10.00	7.53 (1.41)	26
Long EDI Emotional maturity	3247	1.67	10.00	6.56 (1.25)	30
Long EDI Language & cognitive development	3246	0	10.00	9.41 (1.39)	26
Long EDI Communication & general knowledge	3247	0	10.00	7.92 (1.90)	8
SDQ Emotional symptoms	3016	0	10.00	3.89 (2.21)	5
SDQ Conduct problems	3016	0	10.00	3.34 (1.99)	5
SDQ Hyperactivity/inattention	3016	0	10.00	4.44 (1.41)	5

Table 2 (continued)

Instrument/Variable	N	Minimum	Maximum	Mean (SD)	Number of items
SDQ Peer problems	3016	0	9.00	2.29 (1.48)	5
SDQ Pro-social behaviour ^a	3016	0	10.00	2.62 (1.79)	5
SDQ Total Score	3016	1	32.00	13.95 (4.83)	20
Language school test (6 & 7 year olds)	584	0	23.00	12.96 (7.04)	23
Mathematics school test (6 & 7 years olds)	584	0	15.00	9.00 (4.69)	15
Raven's Progressive Matrices (6 & 7 year olds)	584	0	14.00	6.05 (3.54)	14
Language school test (8 & 9 year olds)	2361	0	21.00	14.99 (5.31)	21
Mathematics school test (8 & 9 years olds)	2361	0	25.00	17.29 (6.41)	25
Raven's Progressive Matrices (8 & 9 year olds)	2361	0	17.00	6.89 (3.81)	18

^a SDQ responses are on a 3 point scale with values corresponding to always, sometimes and never. Depending on the response a maximum of 2 points may be awarded per question

^b The Prosocial behaviour scale of the SDQ has been reverse scored such that high scores on this scale indicate a child has poor prosocial behaviour

through to “sometimes disagree”). There was much debate amongst the interviewers - Government staff from the Department of Early Childhood within the MoNE and one experienced fieldworker who had worked on the Indonesian Family Life Survey, among others - about the response options. The expected responses in the Indonesian language would be “can” (bisa) or “can not” (tidak bisa) and the interviewers were unsure about whether it would be possible to generate a range of responses from the respondents. The interviewers were also concerned that the child’s caregivers may demonstrate a strong social desirability response bias when completing the EDI. To help reduce the chance of this bias occurring and to increase the ability to assess reliability of the instruments, the study team added a series of child tasks, fieldworker observations and repeat questions after the pilot testing. Further translations and back translations were undertaken to improve the understanding and interpretability of the ranged response options.

In September 2007 the updated instrumentation suite was piloted in Subang (West Java) with 26 households. As was the case in the previous pilot, respondents struggled with the ranged response categories. It was also noted that respondents found it difficult to provide answers to questions that required some degree of subjectivity in the answer (e.g., How would you rate [*child's name*]'s overall social/emotional development?). Questions such as these required greater guidelines and training for the interviewers.

After making some additional modifications, the EDI was again back translated to English for checking before a final and comprehensive pilot in January 2009 with 100 children across the city of Yogyakarta and 100 children across the rural region of Kebumen. This pilot enabled us to determine any potential ceiling/floor problems, timing of the questionnaire suite and final minor adjustments to the adaptation and translation of all the instruments along with an opportunity to test the fieldworker training and associated fieldworker manuals. Table 4 provides a list of the EDI questions that were modified and the Indonesian version of the question (back translated into English).

Table 3 Descriptive statistics for child indicators

	Time 1 - older cohort		Time 2 – younger cohort	
	Frequency	Percent	Frequency	Percent
Child Indicators				
DCCS (Card Sort) Task				
Unable to complete Stage 2 (colour and shape)	975	34.82	554	21.30
Completed Stage 2 (colour and shape)	1376	49.14	1694	65.13
Completed Stage 3 (colour, shape and border)	449	16.04	353	13.57

As previously noted, at Time 2, the long form of the EDI was used in Indonesia with all modifications implemented for the short form of the EDI maintained. After these three pilots the long EDI was reduced from 104 to 101 items, and the short form EDI was reduced from 47 to 45 items. The items deleted were those that were specific to the school setting, or those that couldn't be satisfactorily translated to Bahasa.

Table 4 List of changes to the EDI questions through the adaptation process

Original EDI Question	EDI Question after adaptation (Back-translated to English)
Too tired and/or sick to do school work	Does your child often complain that he/she is too tired/ill so he/she cannot play/do everyday activities
Hungry	Does your child always eat breakfast
Independent in washroom habits most of the time	Can your child use toilet on his/her own (including taking off his/her pants/skirt and put it on again, passing water/moving his/her bowels, clean after his/herself and wash hands
Is well-co-ordinated (i.e., moves without tripping over things)	Has your child's body movement been well coordinated for example walking/running without bumping into or stumbling on something
Ability to manipulate objects	Is your child able to move, transport, hold, use objects including small objects such as beads
Overall physical development	How do you evaluate the physical development/growth condition of your child compared to other children of his/her age
Ability to use language effectively in English	Is your child able to use appropriate and correct words in Bahasa/local language when he/she has a question or asks for something that he/she wants
Interested in reading	Is your child interested in reading materials and reading, for example wants to know about the meaning of the reading material
Is able to attach sounds to letters	Can your child connect/relate sound/pronunciation with letters
Overall social development	What is the condition of social and emotional growth of your child compared to other children his/her age
Bullies or is mean to others	Does your child often annoy, force, dominate other children that are smaller or weaker for example pushing another child, taking another child's toy forcefully, alienates his/her friends when they play together

3.3 Concurrent Construct Validity

The concurrent construct validity of the EDI in Indonesia can be assessed by looking at how well scores on the EDI's domains are associated with scores on other instruments designed to measure similar constructs, in this case, the SDQ and the DCCS. All instruments are collected at the same time (i.e., concurrently). Concurrent data collection using these measures are available for the older cohort at Time 1 and for the younger cohort at Time 2. Table 5 presents these correlations.

The patterns of correlations between caregiver-rated EDI scores and SDQ scores were similar for the younger and older cohorts. Negative correlations were expected as high scores on the SDQ indicate behavioural problems while high scores on the EDI indicate that a child is developing well. For both the younger and older cohorts, the caregiver-rated Emotional maturity and, to a lesser extent, the Social competence domains of the EDI were moderately to highly correlated with each of the SDQ scales and the SDQ Total difficulties score. Correlations between the Language and cognitive skills and Communication skills and general knowledge domains of the EDI with the scales of the SDQ were generally weak. However there were some moderate correlations between these two EDI domains and the SDQ Prosocial behaviour scale and, in the case of the younger cohort, the SDQ hyperactivity inattention scale.

The magnitude of the correlations between the caregiver-rated EDI domains and the DCCS task were consistent across cohorts. These correlations, while weak, were strongest for the Social competence, Language and cognitive skills, and Communications skills and general knowledge domains of the EDI.

The magnitude of the correlations between teacher-rated EDI domain scores and both the SDQ and DCCS were notably smaller than the corresponding caregiver-rated EDI domains. The majority of the associations were negative, as expected, but the strongest correlations were between the Language and cognitive skills domain and the SDQ, rather than the Social competence and Emotional maturity domains.

3.4 Predictive Validity

The short version of the EDI was completed by caregivers for children in the older cohort at Time 1 (approximately 4 years old) and at Time 2, these same children sat school-based tests assessing their language, mathematics and cognitive performance and their caregivers completed the SDQ. These data allow us to assess the predictive validity of the EDI.

The predictive ability of the EDI's domains was assessed by a regression analysis similar to that of Forget-Dubois et al. (2007). The model aims to predict school outcomes as assessed by language, mathematics and cognitive ability at ages 6 and 7, and then ages 8 and 9 (as different tests were used for the two age ranges). The Language and cognitive skills domain of the EDI was the only consistent significant predictor of all later school test performance with a one unit increase in scores on this domain associated with a 1.73 point higher age 8 and 9 mathematics score. The Social competence and Communication skills and general knowledge domains were significant predictors of later mathematics performance

Table 5 Correlations (Spearman's ρ) between EDI domains and other child development measures assessed concurrently

	SDQ ^a Emotional Symptoms	SDQ Conduct Problems	SDQ Hyperactivity/inattention	SDQ Peer problems	SDQ Prosocial Behaviour ^b	SDQ Total Difficulties	DCCS
Time 1 – older cohort (Caregiver short EDI)	(N = 3251)	(N = 3251)	(N = 3250)	(N = 3250)	(N = 3241)	(N = 3249)	(N = 2797)
Physical health & well-being	-.26***	-.11***	-.08***	-.15***	-.04*	-.24***	0.06**
Social competence	-.16***	-.17***	-.25***	-.21***	-.51***	-.29***	0.15***
Emotional maturity	-.45***	-.53***	-.28***	-.36***	-.34***	-.63***	0.08***
Language & cognitive skills	-.07**	-.09***	-.17***	-.07**	-.27***	-.14***	0.18***
Communication skills & general knowledge	-.09***	-.03	-.06**	-.11***	-.23***	-.11***	0.12***
Time 2 – younger cohort (Caregiver long EDI)	(N = 2771)	(N = 2771)	(N = 2771)	(N = 2770)	(N = 2770)	(N = 2770)	(N = 2590)
Physical health & well-being	-.14***	-.11***	-.21***	-.11***	-.15***	-.21***	.06**
Social competence	-.16***	-.15***	-.35***	-.17***	-.51***	-.29***	.18***
Emotional maturity	-.49***	-.51***	-.31***	-.31***	-.22***	-.63***	.10***
Language & cognitive skills	-.15***	-.06**	-.25***	-.14***	-.28***	-.21***	.24***
Communication skills & general knowledge	.02	.08***	-.22***	-.07**	-.31***	-.07**	.14***
Time 2 - younger cohort (Teacher long EDI)	(N = 1244)	(N = 1244)	(N = 1244)	(N = 1244)	(N = 1244)	(N = 1244)	(N = 1188)
Physical health & well-being	-.05	-.01	-.07*	.02	-.02	-.04	.08**
Social competence	-.08**	-.06*	-.07*	-.07*	-.06*	-.10**	.10**
Emotional maturity	-.09**	-.06*	-.10**	-.07*	-.05	-.11**	.07*
Language & cognitive skills	-.12***	-.08**	-.14***	-.09**	-.09**	-.16***	.16***
Communication skills & general knowledge	-.09**	-.01	-.06*	-.08**	-.09**	-.08**	.13***

^a SDQ was always caregiver-rated

^b The Prosocial behaviour scale of the SDQ has been reverse scored such that high scores on this scale indicate a child has poor prosocial behaviour

*** $p < .0001$, ** $p < .01$, * $p < .05$

(Social competence for ages 8 and 9 performance only). The Social competence and Communication skills and general knowledge domains were significant predictors of performance on the test of cognitive ability at ages 8 and 9.

Following Forget-Dubois et al. (2007) a second series of regressions were carried out (Model 2, see Table 6) in which the five EDI domains were simultaneously entered into the regression along with potential confounders - age, gender and wealth. Results for the five domains were generally consistent with Model 1, with the Language and cognitive skills domain again the only consistent predictor of later school test performance and remaining a predictor even after controlling for age, gender and wealth. The Communication skills and general knowledge domain did not significantly predict later school test performance.

Results of regression analyses predicting SDQ scale scores at 8 years old from EDI scores at 4 years old followed the same pattern as the concurrent correlations and are presented in Table 7. The Emotional maturity domain of the EDI was the most consistent predictor of performance, though the magnitudes of the effects were small with, for example, a one point increase in this domain resulting in a 0.16 point decrease in the Conduct problems scale of the SDQ.

3.5 Inter-rater Reliability

For the younger cohort, the EDI was completed by a child's primary caregiver as well as their teacher at Time 2. Comparing caregiver and teacher EDI scores for the same children allows inter-rater reliability to be assessed via correlations. Table 8 shows that the EDI scores assigned by teachers were generally higher than those assigned by caregivers (the exception being the Communication skills and general knowledge domain). The inter-rater reliability, as assessed by the correlation between caregiver and teacher scores, was best for the Language and cognitive skills domain of the EDI, however the results are very weak, and lower than have been found in other countries.

3.6 Internal Consistency Reliability

The internal consistency reliability of each of the five EDI domains was calculated using Cronbach's alpha. The internal consistency reliability was calculated for both the short EDI (completed by the older cohort's caregivers at Time 1) and the long EDI (completed by the younger cohort's caregivers and teachers at Time 2) (see Table 9). For comparison sake, we also calculated the Cronbach alpha for the SDQ scales (see Table 10). A Cronbach's alpha value of around 0.7 is generally considered to indicate acceptable reliability of a scale (Hair et al. 1998). All domains but the Physical health and well-being domain had values of Cronbach's alpha in excess of or close to 0.7. The long version of the Physical health and well-being domain had significantly better internal consistency than the short version, suggesting that the addition of the extra questions on this domain make it more homogenous. Please note that Confirmatory Factor Analyses have been reported previously as the subject of a previous paper comparing the equivalence of constructs of the EDI across Indonesia and the Philippines (Duku et al. 2015).

Table 6 Regression analysis predicting school language, mathematics and cognitive performance from the EDI domains (standardized coefficients)

	Language				Mathematics				Raven's Progressive Matrices			
	Ages 6 & 7		Ages 8 & 9		Ages 6 & 7		Ages 8 & 9		Ages 6 & 7		Ages 8 & 9	
	Model 1 (N = 583)	Model 2 (N = 2359)	Model 1 (N = 2336)	Model 2 (N = 583)	Model 1 (N = 583)	Model 2 (N = 2336)	Model 1 (N = 2359)	Model 2 (N = 583)	Model 1 (N = 583)	Model 2 (N = 2359)	Model 1 (N = 583)	Model 2 (N = 2336)
Physical health & well-being	.01	.03	.02	.01	.02	.04	.01	-.01	-.06	-.04	.02	.01
Social competence	-.03	-.04	.08**	.05*	.04	.02	.05*	.02	.06	.04	.03	.01
Emotional maturity	.10*	.09*	.07***	.06**	.08	.07	.05	.04	.08	.08	.03	.03
Language & cognitive skills	.27***	.22***	.25***	.19***	.23***	.18***	.27***	.21***	.17***	.13**	.19***	.16***
Communication skills & general knowledge	.08	.08	.04	.04*	.05	.05	.03	.04	.05	.05	.002	.01
Child age				.07***		.04		.09***		-.04		.06**
Gender		.03		.12***		.04		.08***		.01		-.01
Wealth		.24***		.22***		.24***		.24***		.21***		.14***

*** $p < .0001$, ** $p < .01$, * $p < .05$

Table 7 Regression analysis predicting SDQ scale scores (unstandardized coefficients). Model 1 *N* = 3013, Model 2 *N* = 3011

	SDQ Emotional Symptoms		SDQ Conduct Problems		SDQ Hyperactivity/inattention		SDQ Peer problems		SDQ Prosocial Behaviour ^a		SDQ Total Difficulties	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Physical health & well-being	-.14***	-.14***	-.03	-.04	-.01	-.01	-.11***	-.11***	.003	.001	-.30***	-.29***
Social competence	.05	.06	-.02	.01	.003	.01	-.001	-.001	-.17***	-.15***	.06	.08
Emotional maturity	-.14***	-.15***	-.22***	-.21***	-.10***	-.09***	-.03	-.02	-.06**	-.05*	-.49***	-.46***
Language & cognitive skills	-.04	-.04	-.001	.002	-.06***	-.05**	-.03	-.01	-.10***	-.07***	-.13**	-.09
Communication skills & general knowledge	-.03	-.04	.02	.02	-.04	-.04	-.03	-.03	.01	.003	-.08	-.08
Child age		-.01		.09		-.23***		-.14*		.13		-.27
Gender		.09*		-.18***		-.09**		-.05		-.12***		-.23**
Wealth		-.12*		.05		-.09**		-.12***		-.22***		-.28**

*** *p* < .0001, ** *p* < .01, * *p* < .05

Table 8 Caregiver and teacher inter-rater reliability (younger cohort) – Means (SDs) and correlations (Spearman's ρ)

	N	Long EDI (Caregiver) Mean (SD)	Long EDI (Teacher) Mean (SD)	Correlation (ρ)
Physical health & well-being	1245	7.95 (1.35)	8.22 (1.28)	.08**
Social competence	1243	6.75 (1.41)	6.94 (1.60)	.10**
Emotional maturity	1242	5.91 (1.27)	6.40 (1.36)	.11**
Language & cognitive skills	1244	5.17 (2.45)	6.98 (2.38)	.42***
Communication skills & general knowledge	1245	6.91 (2.06)	6.17 (1.89)	.17***

*** $p < .0001$, ** $p < .01$, * $p < .05$

Table 9 Internal consistency reliability of the domains of the short and long versions of the EDI (Cronbach's α)

	Time 1 (older cohort)	Time 2 (younger cohort)	
	Short EDI (Caregiver)	Long EDI (Caregiver)	Long EDI (Teacher)
Physical health & well-being	.19	.57	.63
Social competence	.58	.90	.94
Emotional maturity	.66	.84	.88
Language & cognitive skills	.74	.96	.94
Communication skills & general knowledge	.70	.89	.88

Table 10 Internal consistency reliability of the domains of the SDQ scales (Cronbach's α)

	Time 1 (older cohort) Caregiver response
Emotional problems scale	.57
Conduct problems scale	.38
Hyperactivity scale	.42
Peer problems scale	.27
Prosocial scale	.59
Total score	.72

4 Discussion

To our knowledge, this is the first paper to report validity evidence of the short and standard versions of the EDI, and indeed any measure of child development, in Indonesia. Both the long and short versions of the EDI have been translated and used in an evaluation of ECED services in Indonesia. Content validity had been demonstrated over a number of years with several revisions to the instrument done over this time, in particular the response options and the wording of the translation.

The concurrent construct validity of the EDI in Indonesia is comparable to its validity in other countries. As is the case in Canada and Australia, in Indonesia the caregiver-rated EDI's Social competence and Emotional maturity domains were moderately correlated with the domains of the SDQ. Most associations found were of similar strength to those identified in previous work (Janus et al. 2007), however the associations were notably weaker in this study for teacher-rated Social competence and Emotional maturity. Given that the SDQ was completed by a child's caregiver, the smaller correlations with the teacher-rated EDI compared to the caregiver-rated EDI may be reflective of the fact that in the latter case, the respondent is the same person for both the EDI and SDQ, consistent with the work of Brinkman et al. (2007).

Previous research has also found that the Language and cognitive skills domain of the EDI is moderately correlated with other tests of cognitive ability including the PPVT, the Who am I? test and the HKECDS (Ip et al. 2013; Janus et al. 2011; Janus and Offord 2007). The DCCS was the measure of cognitive ability used in Indonesia. Consistent with previous work, performance on the DCCS task correlated most strongly with a child's score on the Language and cognitive skills domain of the EDI, however the magnitude of this correlation was small. The DCCS task itself was relatively difficult to implement in the Indonesian context with most people unfamiliar with this type of task. The weak correlations may therefore also reflect the validity of the DCCS task in Indonesia, as well as the EDI.

Previous research on the predictive validity of the EDI has found that a child's scores on particular domains of the EDI are good predictors of their teacher-reported school achievement 1 year later (Forget-Dubois et al. 2007), their Grade 2 visual-motor integration ability, SDQ scores, and learning aptitude (Janus et al. 2007), and their performance on school-based tests of literacy and numeracy performance in Grades 3, 5, and 7 (Brinkman et al. 2013). Predictive validity results obtained from this Indonesian study are generally consistent with these previous findings; however, in all circumstances the magnitudes of the relationships found were weaker.

Out of the five domains of the EDI the Language and cognitive development domain performs the best. In the Indonesian context, a child's score on the Language and cognitive skills domain of the EDI at approximately 4 years old was moderately correlated with their language and mathematics ability at ages 6 and 7 and, more weakly, at ages 8 and 9. The Language and cognitive skills domain of the EDI was also a moderate predictor of a child's later performance on Prosocial behaviour as measured by the SDQ. Further, the Language and cognitive skills domain was still a good predictor of later school assessment after controlling for age, gender and wealth, and indeed, a child's score on this domain was a more powerful predictor of later school performance than any of these demographic characteristics.

For the younger cohort at the Time 2 data collection, both a child's teacher and their primary caregiver completed the EDI allowing us to assess the inter-rater reliability of the EDI in Indonesia. The magnitude of the correlations between these informants was small and the relationship weaker than the inter-rater reliability of the EDI in other countries. In Indonesia, the inter-rater reliability ranged from 0.10 to 0.41 compared with correlations between Canadian parents and teachers in the range of 0.36 to 0.64 (Janus and Offord 2007). Inter-rater reliability between teachers is generally found to be higher than that between teachers and parents (Janus et al. 2007) and, given their experience with many children, teachers are generally thought to be better judges of a child's abilities than are a child's caregivers (e.g., Sommer et al. 2008). However, in this study, assessment of the predictive ability of the EDI is only possible for caregiver-rated EDI and, despite approximately 4 years elapsing, these caregiver-rated EDI domain scores are predictive of a child's later outcomes, particularly school-based assessment.

The internal consistency reliability of the EDI in Indonesia was comparable with Canada, Australia, Jamaica, USA, Scotland and China (Forer and Zumbo 2011; Forget-Dubois et al. 2007; Ip et al. 2013; Janus et al. 2007, 2011; Janus and Offord 2007; Woolfson et al. 2013). In each of these locations, the internal consistency reliability was lowest for the Physical health and wellbeing domain (the lowest in these countries was .64 in Jamaica, which is very similar to the teacher-completed long EDI in Indonesia). The internal consistency for the Physical health and well-being domain in the short version of the EDI was very poor suggesting that it is very important when assessing this domain to include the additional items that are present in the long version of this domain. The additional items in the long version of the Physical health and well-being domain were related to whether the child was appropriately dressed for daily activities, whether they were often late, were proficient at holding a pen, their level of energy and their personal hygiene. We are unable to know if the long version of the EDI had indeed been collected at Time 1 for the older cohort whether we would have found stronger results for the predictive validity assessment.

This study has shown that both the long and short versions of the EDI show weak to moderate reliability and validity in the Indonesian context. However, given that the internal consistency reliability, particularly for the Physical health and well-being domain, is superior for the long version, it would seem prudent to use the long version of the EDI and at least the long version of this domain, where possible.

In light of these findings, there are two extreme approaches one can take; one could take the view that any adaptation of a western instrument to a very different culture and context is culturally imperialist and should be avoided, others may take the view that to enable any degree of international comparability, instruments should remain as consistent with their original form as possible. This latter group may well assert that translations with minimal to no adaptation are the best recourse. In the present example in Indonesia the process of adaptation to maintain test equivalence across culture without bias was a challenge. This paper explored the processes we undertook and presents various aspects of reliability and validity for the EDI in Indonesia. On balance, it would seem that the child development outcomes as ascertained by the EDI in Indonesia are valid and reliable. However, these results are not as strong as we had initially hoped despite significant efforts. These efforts took the form of a series of initial workshops around the relevance of the content of the EDI, the inclusion of local

Indonesian researchers and multiple pilots and translations and back translations of the Instrument over years.

Our findings may be of interest to others working in early child development and education in non-western settings looking for international comparability. The development of reliable instruments that can be used to compare the impact of programs in different countries evaluating early childhood education and development initiatives remains a challenge for the field. Taken at the two extremes it would seem to be more important to determine if programs are making an impact for children on outcomes that are meaningful locally, with international comparability relegated to a secondary aim.

Acknowledgments Our collaboration has been partially supported by an Australian Government AusAID Development Research Awards Scheme Grant (ADRA0800261). Data collection was partially funded by the Government of the Kingdom of the Netherlands through the Dutch Education Support Program (DESP) Trust Fund (TF057272) which provides support to the Government of Indonesia through the World Bank for the purpose of developing policies, studies, and programs that help the Government achieve its education strategic plan. The findings, interpretations, and conclusions expressed in this paper, do not necessarily reflect the views of the Government of the Kingdom of the Netherlands or the Government of Indonesia.

We would particularly like to acknowledge the AusAID Education Thematic Group and Network Members for their support and interest in the translation of these results into evidence-based policy and practice. In carrying out this project we have worked closely with our colleagues at the World Bank, including Amanda Beatty, Hafid I. Alatas, Joppe de Ree, Titie Hadiyati, Djoko Hartono, Dedy Junaedi, Mayla Safuro, Mulyana and Rosfita Roesli.

References

- Alatas, H., Brinkman, S., Chang, M. C., Hadiyati, T., Hartono, D., Hasan, A., & Roesli, R. (2013). Early childhood education and development services in Indonesia. In D. Suryadarma & G. W. Jones (Eds.), *Education in Indonesia*. Singapore: ISEAS Publishing.
- Bernard van Leer Foundation. (2006). *A guide to general comment 7: implementing child rights in early childhood*. The Hague: United Nations Committee on the Rights of the Child, United Nations Children's Fund and Bernard van Leer Foundation.
- Brinkman, S. (2009, 16–18 November). *The Impact and Reach of the EDI Around the World*. Paper presented at the The Early Development Imperative: A Pan-Canadian Conference on Population Level Measurement of Children's Development, Winnipeg, Canada.
- Brinkman, S., Silburn, S., Lawrence, D., Goldfeld, S., Sayers, M., & Oberklaid, F. (2007). Investigating the validity of the Australian Early Development Index. *Early Education and Development, 18*(3), 427–451. doi:10.1080/10409280701610812.
- Brinkman, S., Gregory, T., Harris, J., Hart, B., Blackmore, S., & Janus, M. (2013). Associations between the early development instrument at age 5 and reading and numeracy skills at ages 8, 10 and 12: a prospective linked data study. *Child Indicators Research, 6*(4), 695–708. doi:10.1007/s12187-013-9189-3.
- Carlson, S. M. (2005). Developmentally sensitive measures of executive functioning in preschool children. *Developmental Neuropsychology, 28*(2), 595–616. doi:10.1207/s15326942dn2802_3.
- Central Board of Statistics of Indonesia. (2010). National Socioeconomic Survey 2010 (SUSENAS 2010). Jakarta, Indonesia.
- Duku, E., Janus, M., Brinkman, S. (2015). Investigation of the cross-national equivalence of a measurement of early child development. *Child Indicators Research, 8*(2): 471–489. doi:10.1007/s12187-014-9249-3.
- Forer, B., & Zumbo, B. D. (2011). Validation of multilevel constructs: validation methods and empirical findings for the EDI. *Social Indicators Research, 103*(2), 231–265. doi:10.1007/s11205-011-9844-3.
- Forget-Dubois, N., Lemelin, J. P., Boivin, M., Dionne, G., Séguin, J. R., Vitaro, F., & Tremblay, R. E. (2007). Predicting early school achievement with the EDI: a longitudinal population-based study. *Early Education and Development, 18*(3), 405–426. doi:10.1080/10409280701610796.
- Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*(4), 483–527. doi:10.1016/0885-2014(95)90024-1.

- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of American Academy of Child and Adolescent Psychiatry*, 40, 1337–1345.
- Goodman, R. (2005). Strengths and Difficulties Questionnaire - Indonesian translation. Retrieved 14/12/2015, from <http://www.sdqinfo.com/py/sdqinfo/b3.py?language=Indonesian>.
- Goodman, R., Renfrew, D., & Mulick, M. (2000). Predicting type of psychiatric disorder from Strength and Difficulties Questionnaires (SDQ) scores in child mental health clinics in London and Dhaka. *European Child and Adolescent Psychiatry*, 9, 129–134.
- Hair, J. E., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate Data Analysis* (5th ed.). New Jersey: Prentice-Hall.
- Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2005). *Adapting educational and psychological tests for cross cultural assessment*. Mahwah: Lawrence Erlbaum Associates Inc.
- Hasan, A., Hyson, M., & Chang, M. C. E. (2013). *Early Childhood Education and Development in Poor Villages of Indonesia*. Washington DC: International Bank for Reconstruction and Development / The World Bank.
- Hawes, D. J., & Dadds, M. R. (2004). Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. *Australian and New Zealand Journal of Psychiatry*, 38(8), 644–651.
- Hayes, L. (2007). Problem behaviours in early primary school children: Australian normative data using the strengths and difficulties questionnaire. *Australian and New Zealand Journal of Psychiatry*, 41, 231–238.
- Herdman, M., Fox-Rushby, J., & Badia, X. (1998). A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Quality of Life Research*, 7, 323–335.
- How to use the EDI. (2015). Retrieved 14/12/2015, from <https://edi.offordcentre.com/researchers/how-to-use-the-edi/>.
- International Test Commission (ITC). (2000). International Guidelines for Test Use. Retrieved 11/8/2010, from <http://www.intestcom.org/guidelines/index.php>.
- Ip, P., Li, S. L., Rao, N., Ng, S. S. N., Lau, W. W. S., & Chow, C. B. (2013). Validation study of the Chinese Early Development Instrument (CEDI). *BMC Pediatrics*, 13(1), 146. doi:10.1186/1471-2431-13-146.
- Janus, M., & Duku, E. K. (2005). Development of the Short Early Development Instrument (S-EDI). Report for the World Bank.
- Janus, M., & Offord, D. (2007). Development and psychometric properties of the Early Development Instrument (EDI): a measure of children's school readiness. *Canadian Journal of Behavioural Science*, 39, 1–22.
- Janus, M., Brinkman, S., Duku, E., Hertzman, C., Santos, R., & Sayers, M. (2007). *The early development instrument: A population-based measure for communities. A handbook on development, properties and use*. Hamilton: Offord Centre for Child Studies.
- Janus, M., Brinkman, S., & Duku, E. (2011). Validity and psychometric properties of the early development instrument in Canada, Australia, United States and Jamaica. *Social Indicators Research*, 103(2), 283–297. doi:10.1007/s11205-011-9846-1.
- Mize, L., Pambudi, E., Koblinsky, M., Stout, S., Marzoeki, P., Harimurti, P., & Rokx, C. (2010). “.....and then she died”: Indonesian maternal health assessment. In Health Sector Review (Ed.). Jakarta: World Bank.
- Pradhan, M., Brinkman, S. A., Beatty, A., Maika, A., Satriawan, E., de Ree, J., & Hasan, A. (2013). Evaluating a community-based early childhood education and development program in Indonesia: study protocol for a pragmatic cluster randomized controlled trial with supplementary matched control group. *Trials*, 14(1), 259. doi:10.1186/1745-6215-14-259.
- Sabbagh, M., Xu, F., Carlson, S., Moses, L., & Lee, K. (2006). The development of executive functioning and theory-of-mind: a comparison of Chinese and US preschoolers. *Psychological Science*, 17(1), 74–81.
- Sommer, U., Fink, A., & Neubauer, A. C. (2008). Detection of high ability children by teachers and parents: psychometric quality of new rating checklists for the assessment of intellectual, creative and social ability. *Psychology Science Quarterly*, 50(2), 189–205.
- UNICEF. (2009). *The State of the World's Children Special Edition: Celebrating 20 Years of the Convention on the Rights of the Child*. New York: United Nations Children Fund.
- Woolfson, L., Geddes, R., McNicol, S., Booth, J., & Frank, J. (2013). A cross-sectional pilot study of the Scottish early development instrument: a tool for addressing inequality. *BMC Public Health*, 13(1), 1187. doi:10.1186/1471-2458-13-1187.
- World Bank. (2010). *Indonesia Economic Quarterly: Continuity Amidst Volatility*. Jakarta.
- World Bank. (2012). *Targeting Poor and Vulnerable Households in Indonesia*. Jakarta: The World Bank.
- World Bank. (2013). *Slower growth; high risks. Indonesia Economic Quarterly, December 2013*. <http://www.worldbank.org/content/dam/Worldbank/document/EAP/Indonesia/IEQ-Dec13-ENGLISH.pdf>.
- Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): a method of assessing executive function in children. *Nature Protocols*, 1(1), 297–301. doi:10.1038/nprot.2006.46.