

Exploring change over time in the Australian Early Development Census (AEDC): The estimation of a critical difference for a wide range of AEDC indicators

**Tess Gregory and Sally Brinkman** 





# Acknowledgements

We would like to acknowledge Martin Guhn, Barry Forer and Bruno Zumbo for the development of the critical difference methodology and particularly Martin and Barry for being so generous in supporting our adaptation of their methods to the Australian data their comments on earlier drafts of this report. We would also like to thank Eric Duku, Magdalena Janus, and Joanne Schroeder for their contribution and support of a consistent methodology across countries, and their comments on earlier drafts of this report.

# A note on this report

The current report is an updated version of a technical report that was original published in February 2013.

Gregory, T., & Brinkman, S. (2013). Methodological approach to exploring change in the Australian Early Development Index (AEDI): The estimation of a critical differences. Telethon Institute for Child Health Research, Western Australia.

The original report focused on estimating the critical difference for the AEDC indicators that focused on the percentage of children who were "vulnerable". Specifically, the percentage of children who were vulnerable on each of the five AEDC domains, as well as two summary indicators, developmentally vulnerable on 1 or more domains (DV1) and developmentally vulnerable on 2 or more domains (DV2). In 2016, we extended the original work by calculating a critical difference for the percentage of children who were "at risk" and "on track" in each of the five domains. In 2022, this was extended further to calculate the critical difference for a new summary indicator – On Track on five domains (OT5).

In this report, we use the term Australian version of the Early Development Instrument (AvEDI) to refer to the child development instrument that teachers complete, and the term Australian Early Development Census (AEDC) to refer to the program of work to conduct a national census of child development once every three years.







# **Report prepared by**

The Telethon Kids Institute, Adelaide Branch.

# **Suggested citation**

Gregory, T., & Brinkman, S. (2022). Exploring change over time in the Australian Early Development Census (AEDC): The estimation of a critical difference for a wide range of AEDC indicators. Telethon Kids Institute, Adelaide, Australia.

# For more information about this report, please contact

Telethon Kids Institute Level 7, 31 Flinders Street Adelaide, SA 5000 (08) 8207 2026 Email: <u>tess.gregory@telethonkids.org.au</u> or <u>sally.brinkman@telethonkids.org.au</u> www.telethonkids.org.au





### Contents

LIST OF TABLES
LIST OF FIGURES
BACKGROUND7
Canada's experience with using the EDI over time8
CALCULATING THE CRITICAL DIFFERENCE: A FOUR STAGE PROCESS
STAGE ONE: EXPLORE THE AVEDI FACTOR STRUCTURE AND EXTRACT PARAMETERS
Physical Health and Wellbeing domain13
Social Competence domain14
Emotional maturity domain15
Language and cognitive skills (school-based) domain16
Communication skills and general knowledge domain17
STAGE TWO: RUN THE COMPUTER SIMULATION18
Step 1: Generate true scores for the five domains18
Step 2: Generate the observed scores for each of the items
Step 3: Combine the 96-items into the domain scores and compute all categorical indicators 20
Step 4: Examine the distributions for each of the indicators and extract the SD estimate 21
STAGE THREE: ESTIMATE THE CRITICAL DIFFERENCE FOR ALL COMMUNITY SIZES
STAGE FOUR: ESTIMATE THE POWER FUNCTION FOR EACH OF THE INDICATORS
Power functions: % vulnerable
Power functions: % at risk
Power functions: % on track
Power functions: Summary indicators
Power functions: All indicators 40
USING THE CRITICAL DIFFERENCE IN PRACTICE
Using the power functions
Using the Comparative Results Tool41
Using look up tables
SUMMARY OF FINDINGS
REFERENCES







# **List of Tables**

Table 1: Factor loadings for the 12-items from the physical health and wellbeing domain	13
Table 2: Factor loadings for the 24-items from the social competence domain	14
Table 3: Factor loadings for the 26-items from the emotional maturity domain	15
Table 4: Factor loadings for the 26-items from the language and cognitive skills domain	16
Table 5: Factor loadings for the 8-items from the communication skills and general knowledge domain	17
Table 6: Deciles of community size for 2009 AEDC data	18
Table 7: Computer simulation results – Developmental vulnerability on the social competence domain	23
Table 8: Critical differences – Developmental vulnerability on the Social competence domain	24
Table 9: Critical differences for all "vulnerable" based indicators for the AvEDI	25
Table 10: Power functions for each of the 5 developmental domains for each of the 3 categories	40
Table 11: Power functions to calculate the critical difference for the 2 summary indicators	40
Table 12. Look up table of critical differences for five AvEDI domains - On track indicators	43
Table 13. Look up table of critical differences for five AvEDI domains – At risk indicators	44
Table 14. Look up table of critical differences for five AvEDI domains – Vulnerable indicators	45
Table 15. Look up table of critical differences for summary indicators	46





# List of Figures

Figure 1: Distribution for % vulnerable on two AvEDI domains for communities with 35 children	21
Figure 2: Distribution for % vulnerable for communities with 35 children (alternate run)	22
Figure 3: Plot of critical differences x community size – % vulnerable in the Social competence	27
Figure 4: Critical difference for % vulnerable on the Physical Health and Wellbeing domain	29
Figure 5: Critical difference for % vulnerable on the Social Competence domain	29
Figure 6: Critical difference for % vulnerable on the Emotional Maturity domain	30
Figure 7: Critical difference for % vulnerable on the Language and Cognitive Skills domain	30
Figure 8: Critical difference for % vulnerable on the Communication and General Knowledge domain	31
Figure 9: Critical difference for % at risk on the Physical Health and Wellbeing domain	32
Figure 10: Critical difference for % at risk on the Social Competence domain	32
Figure 11: Critical difference for % at risk on the Emotional Maturity domain	33
Figure 12: Critical difference for % at risk on the Language and Cognitive Skills domain	33
Figure 13: Critical difference for % at risk on the Communication and General Knowledge domain	34
Figure 14: Critical difference for % on track on the Physical Health and Wellbeing domain	35
Figure 15: Critical difference for % on track on the Social Competence domain	35
Figure 16: Critical difference for % on track on the Emotional Maturity domain	36
Figure 17: Critical difference for % on track on the Language and Cognitive Skills domain	36
Figure 18: Critical difference for % on track on the Communication and General Knowledge domain	37
Figure 19: Critical difference for Vulnerable on 1 or more domains	38
Figure 20: Critical difference for Vulnerable on 2 or more domains	38
Figure 21: Critical difference for <b>On Track on 5 domains (OT5)</b>	39
Figure 22. Comparative Results Tool (screen shot)	42





### Background

In 2021, the Australian Early Development Census (AEDC) was completed for the fifth time. The Australian Government, State/Territory governments, and communities all wanted to explore changes over time in their results to see whether early child development had changed significantly for different cohorts of children. However, the geographical size of these areas and the numbers of children within the areas of interest varies markedly with over 300,000 children assessed nationally, but less than 30 children assessed within some communities. This report provides the technical details for the methodology used to determine "how big" a difference in the AEDC results between two different time points needs to be, to be considered statistically significant, for groups of children of different sizes. This methodological approach was developed to compare 2009 to 2012 results but can be used to compare results for any two time points (i.e., 2009 to 2021, 2015 to 2021).

Before detailing the methodology for reporting on change in the AEDC, it is important to touch on the appropriateness of talking about statistical significance for census data. The AEDC is, like the name suggests, a census of child development capturing data on the whole population rather than a sample of children. This often leads to the argument that it is sufficient to present point estimates, without any confidence intervals or error measures. However, the Australian version of the Early Development Instrument (AvEDI) is a teacher level assessment, and with all such measures, there is associated measurement error and some level of teacher/class level variation. Some teachers may be harsher or softer in their judgement of a child's abilities, however we also find that by the time the AvEDI is collected (midway into the school year) there has already been some influence of the teacher on their class. The AvEDI is collected midway through the school year so that teachers have known the children for long enough to be able to complete the instrument and long enough for children to have settled into the class routines but not too long so that there has been significant influence of the school/class/teacher on the children. Moreover, other census data tend to be aggregated and presented for sizeable groups, whereas the AEDC data are presented at the community level, where some local communities have only 15 children. All of these factors necessitate the provision of some guidance and cautions to communities about the interpretation of change in AEDC results over time. This particularly applies in areas with small numbers of children and only a couple of teachers completing the data collection.





#### Canada's experience with using the EDI over time

In Canada, the EDI<sup>1</sup> has been used as a population measure of early childhood development since 1999. The EDI data are collected at the provincial level in Canada rather than at the national level. Therefore, different schedules of data collection occur within Ontario compared to British Columbia, and EDI results are reported in different ways in Manitoba compared to Alberta. In most provinces of Canada, the EDI data have been collected at least three times, which provides the opportunity to compare data between two time points but also to explore trends in early childhood development.

Within British Columbia, researchers from the Human Early Learning Partnership (HELP) at the University of British Columbia (UBC) developed a method of comparing EDI results at two points in time for communities (Forer, Guhn, & Zumbo, 2011). This method involves psychometric analysis of the EDI items followed by a series of computer simulations that mimic the process of teachers rating their students on the EDI. These simulations provide information about how much variation in EDI results can be expected based on the community size, measurement error, and teacher/class level variation effects. The method generates a series of 'critical difference' values for naturally occurring groups of different sizes (e.g. communities, school districts), that is, groups with different numbers of children living in them. If the difference between the percentage of children who are developmentally "vulnerable" in the two data collection phases is larger than the 'critical difference', then it is interpreted to represent a statistically significant change.

This critical difference approach primarily draws from a measurement error perspective and from generalizability theory (Cronbach, Rajaratnam, & Gleser, 1963). Given that EDI is completed on almost all children, we don't need to take into account sampling error. However, for example, in those communities where it is known that a large local school didn't participate in the EDI data collection, and we are unsure of how representative the EDI data are then it is still prudent to take into account sampling error.

After discussions between researchers at the Human Early Learning Partnership (HELP) at the University of British Columbia, the Offord Centre for Child Studies at McMaster University (Ontario, Canada) and the Telethon Kids Institute (Perth, Australia), a decision was made to use one consistent method for examining change in the EDI. Within Australia, there was a desire to generate critical difference estimates for all five indicators for the developmental domains, and the two summary indicators (1) vulnerable on 1 or more domains, and (2) vulnerable on 2 or more domains.

<sup>&</sup>lt;sup>1</sup> The Early Development Instrument (EDI) is a teacher completed measure of child development that was developed by Dan Offord and Magdalena Janus in Canada. The EDI was adapted for use in Australia, and the adapted instrument is referred to as the Australian version of the Early Development Instrument (AvEDI). The program of work to collect child development information across Australia is referred to as the Australian Early Development Census (AEDC).





In addition to exploring changes in the percentage of children who were developmentally "vulnerable" over time, some AEDC data users were also interested in comparing the percentage of children who were "at risk" and "on track" on each of the five domains over time. As such, in 2015, the critical difference methodology was extended to estimate a critical difference value for each of the five domains on the "at risk" and "on track" categories. In 2021, the methodology was extended further to explore the critical difference for a new summary indicator – On Track on 5 domains (OT5).

In the next section, we will provide a detailed explanation of the methodological approach to calculating the critical difference values and the resultant critical different values for the Australian version of the Early Development Instrument (AvEDI).





# Calculating the critical difference: A four stage process

The process of calculating critical difference values involves four key stages of analysis. These stages are described briefly below. The results of applying the methodology to the AvEDI data are presented in the subsequent section.

**Stage one** involves an exploration of the factor structure of the domains of the AvEDI. We run five exploratory factor analyses<sup>2</sup>, one for each of the domains, and estimate the factor loadings<sup>3</sup> and residual errors for each of the 96-items in the AVEDI. In addition, we explore the proportion of children who fall into each of the categories (e.g. yes 86%, no 14%, for a binary<sup>4</sup> item) for each item, and the z-score<sup>5</sup> that this would correspond to if the categorical variable was a continuous, normally distributed variable (e.g. z = -1.08). This process was completed twice, once for each of the first two waves of AvEDI data.

**Stage two** uses these parameters in the computer simulation to mimic the process of 1,000 different teachers scoring the AvEDI for the same set of children. We generate an observed score for each of the 96 items in the AvEDI. Once generated, these item scores are combined into the five continuous domain scores, the established cut-offs are applied to classify each child as to whether they are vulnerable (or not), at risk (or not) and on track (or not) on each the five domains, and whether they are vulnerable on 1 or more domains (or not), and vulnerable on 2 or more domains (or not), and on track of five domains (or not). Finally, we collate the results for the 1,000 replications, and for each of the indicators (e.g. % vulnerable on the social competence domain), we examine the distribution and record the mean and standard deviation. Stage two is repeated multiple times under different conditions. Specifically, it is repeated for two different waves of AvEDI data (2009 and 2012) and for a series of specific community sizes. The standard deviation estimates from these runs of the simulation are fed into the stage three calculations.

<sup>&</sup>lt;sup>2</sup> **Factor analysis** is a statistical technique that is used to reduce a large number of variables (or items) into a smaller number of unobserved variables called factors. For example, the 96-items of the AvEDI can be reduced to a smaller number of factors, that correspond to the five domains of the AvEDI.

<sup>&</sup>lt;sup>3</sup> **Factor loadings** denote the strength and direction of association between an individual item and the unobserved variable (i.e., factor). Factor loadings can take values from -1 to 1, like a correlation coefficient. A factor loading of 0.80 suggests that item has a strong, positive association with the underlying, latent factor.

<sup>&</sup>lt;sup>4</sup> Binary variables or binary item can take just two possible values, such as Yes/No or 0/1.

<sup>&</sup>lt;sup>5</sup> **Z-score** is a numerical measurement that describes a value's relationship to the mean of a group of values. Z-score is measured in terms of standard deviations from the mean. If a Z-score is 0, it indicates that the data point's score is identical to the mean score. A z-score of 1, indicates that the data point's score is 1 equivalent to one standard deviation above the mean.





**Stage three** takes the standard deviation estimates from stage two and uses them to estimate the critical differences. For a specific community size and specific AvEDI indicator, we take the standard deviation<sup>6</sup> from Wave 1 (2009) and the standard deviation from Wave 2 (2012) and feed these into a formula to calculate the critical difference. At the end of this stage, we have a table of critical difference values for each of the indicators for a series of different sized communities. For example, if you have a community with 100 children and you want to look at the percentage of children who are vulnerable on the Emotional maturity domain, the critical difference value would be 3.5 percentage points. That is, a community with 100 children would need to see a shift in the level of vulnerability on the emotional maturity domain of 3.5 percentage points or greater to be statistically significant.

**Step four** plots the critical difference values (x-axis) against the community size (y-axis) and examines the curve. In all cases, we find that as the community size increases, the critical difference decreases. A visual inspection of the curve shows that it follows a power function<sup>7</sup>. The exact format of this power function is calculated so that the relationship between the critical difference and community size can be described by a single formula. This formula will allow an individual to enter their community size at each time point and determine the specific critical difference. As there are differing levels of measurement error associated with each of the AvEDI domains and thus there is an individual power functions) and one for each of the three summary indicators (Vulnerable on 1 or more domains, Vulnerable on 2 or more domains, and On Track of 5 domains).

<sup>&</sup>lt;sup>6</sup> Standard deviation is a measure of the amount of variation in a set of scores, specifically the degree to which a set of scores are spread out around the mean score.

<sup>&</sup>lt;sup>7</sup> A **power function** is an equation with a single term that is the product of (1) a real number and (2) a variable raised to a fixed real number. For example,  $f(x) = 56x^3$ 





# Stage one: Explore the AvEDI factor structure and extract parameters

The exploratory factor analyses were run on the first two waves of AvEDI data (2009 and 2012), separately. For each wave, five exploratory factor analyses were run, one for each of the five domains of the AvEDI. For instance, a one factor model was run which specified a latent factor<sup>8</sup>, defined by the 12 items that measure the Physical health and wellbeing domain. Analyses were run in MPlus Version 4.2 (Muthen & Muthen, 1998-2006), with all AvEDI items defined as categorical variables<sup>9</sup>, and the estimator was weighted least squares (WLSM).

The factor analysis provides three pieces of information that go into the computer syntax for the simulations that follow. These are:

- 1) Factor loadings<sup>10</sup> of each item on the domain score
- 2) Residual variances<sup>11</sup> of each item on the domain score
- 3) Threshold values for each item

The factor loadings and residual variance estimates are inversely related, that is, as the factor loading increases the residual variance decreases. Therefore, it is not useful to present both sets of parameters here. Given that the factor loadings are more interpretable, the factor loadings for the 2009 and 2012 AvEDI data are presented in Table 1 to Table 5 below. The threshold values will be discussed after the factor loadings.

<sup>&</sup>lt;sup>8</sup> A **latent factor** is a variable that is not directly observed or measured but is defined by scores on a set of observed variables. For example, in the AvEDI, social competence is a latent factor that is not measured directly but is defined by a child's scores on each of the 24-item that are included in the social competence domain.

<sup>&</sup>lt;sup>9</sup> A categorical variable can take a limited, fixed number of possible values. For example, (1) strongly agree, (2) agree,
(3) disagree, and (4) strongly disagree

<sup>&</sup>lt;sup>10</sup> **Factor loadings** denote the strength and direction of association between an individual item and the unobserved variable (i.e., factor). Factor loadings can take values from -1 to 1, like a correlation coefficient. A factor loading of 0.80 suggests that item has a strong, positive association with the underlying, latent factor.

<sup>&</sup>lt;sup>11</sup> **Residual variance** is the portion of variance in each item that is not related to the latent factor (i.e., unique variance). An item with a high factor loading will have a low residual variance.





#### Physical Health and Wellbeing domain

The items that load most highly on the physical health and wellbeing domain are the items assessing gross and fine motor skills with factor loadings between 0.67 and 0.95. Items assessing other aspects of physical health and wellbeing tend to have much smaller loadings (0.39 to 0.56) suggesting they contribute less to the construct. This suggests that the two sets of items may measure slightly different theoretical constructs. The factor loadings are consistent across the two waves of data, suggesting the relationship between items and the physical health and wellbeing domain is stable over time.

Item code	Item description	2009 AvEDI data	2012 AvEDI data
A9	manipulate objects	.946	.949
A12	overall physical development	.933	.949
A10	climb stairs	.894	.916
A8	holding pen etc	.884	.897
A11	level of energy	.847	.949
A13	daily personal hygiene	.766	.804
A7	well co-ordinated	.666	.675
A3	too tired and/or too sick	.554	.564
A6	hand preference established	.524	.520
A4	hungry	.517	.580
A2	over/under dressed	.500	.534
A5	independent toileting	.414	.394

Table 1: Factor loadings for the 12-items from the physical health and wellbeing domain





#### Social Competence domain

All of the items from the social competence domain load together well, with the factor loadings varying from 0.72 to 0.94. This domain is cohesive and the items correlate well. The factor loadings are very similar for the 2009 data and the 2012 data.

Item code	Item description	2009 AvEDI data	2012 AvEDI data
C9	Respect for other children	.934	.939
C6	Respects property of others	.930	.937
C8	Respect for adults	.930	.936
C5	Follows rules and instructions	.922	.923
C3	Cooperates with other children	.916	.917
C2	Able to along with peers	.903	.897
C7	Self-control	.900	.902
C10	Accepts responsibility for actions	.890	.896
C14	Care of school materials	.881	.890
C1	Overall emotional/social development	.853	.849
C22	Follow class routines without reminders	.854	.859
C11	Listens attentively	.851	.863
C13	Works independently	.831	.840
C4	Play with various children	.850	.856
C21	Follow one-step instructions	.838	.846
C23	Adjust to changes in routine	.815	.825
C19	Eager to play with /read new book	.798	.830
C12	Completes work on time	.796	.823
C18	Eager to play new game	.762	.795
C15	Works neatly and carefully	.757	.771
C25	Tolerance towards others	.754	.772
C20	Solve day-to-day problems by self	.750	.759
C17	Eager to play with new toy	.726	.768
C16	Curious about the world	.722	.764

Table 2: Factor loadings for the 24-items from the social competence domain



#### Emotional maturity domain

Most of the items from the Emotional maturity domain load together well. Most of the items have factor loadings between 0.70 and 0.95. However, several items have factor loadings between 0.42 and 0.70 suggesting that this factor is not as cohesive as the Social competence factor. The factor loadings are consistent across the two waves of data, suggesting the relationship between items and the Emotional maturity domain is stable over time.

Item code	Item description	2009 AvEDI data	2012 AvEDI data
C33	Helps other children who are feeling sick	.951	.956
C30	Comforts child who is crying or upset	.949	.951
C40	Distractible	.889	.894
C31	Help pick up objects dropped by others	.882	.892
C39	Can't sit still, is restless	.876	.877
C45	Cannot settle to anything	.868	.890
C26	Tries to help someone who is hurt	.868	.878
C27	Volunteers help to clean someone else's mess	.868	.875
C29	Help other children with task difficulty	.863	.882
C41	Disobedient	.831	.835
C46	Inattentive	.831	.837
C36	Kicks, bites, hits other children/ adults	.829	.826
C43	Impulsive, acts without thinking	.816	.831
C28	If there is a quarrel, will try to stop	.812	.828
C32	Invite others to join game	.811	.844
C34	Gets into physical fights	.807	.806
C44	Difficulty waiting turns in games/groups	.792	.807
C35	Bullies or is mean to others	.733	.736
C37	Takes things that do not belong to him/her	.722	.730
C38	Laughs at other children's discomfort	.702	.716
C42	Has temper tantrums	.648	.655
C51	Incapable of making decisions	.557	.557
C47	Seems unhappy, sad or depressed	.545	.555
C48	Appears worried	.447	.456
C50	Nervous, highly strung or tense	.444	.448
C49	Cries a lot	.425	.424

Table 3: Factor loadings for the 26-items from the emotional maturity domain





#### Language and cognitive skills (school-based) domain

The items from the language and cognitive skills domain load together well. In general, this domain is cohesive and the items correlate well together. The factor loadings are very consistent across the two waves of data, suggesting the relationship between the items and construct is stable over time.

Item code	Item description	2009 AvEDI data	2012 AvEDI data
B17	Able to read simple sentences	.929	.921
B15	Able to read simple words	.927	.928
B22	Able to write simple words	.898	.907
B11	Identify some letters of alphabet	.889	.899
B12	Able to attach sounds to letters	.878	.880
B9	Generally interested in books	.867	.879
B23	Able to write a simple sentence	.867	.870
B26	Interested in games involving numbers	.864	.875
B25	Interested in mathematics	.858	.866
B16	Able to read complex words	.844	.815
B10	Interested in reading	.833	.854
B30	Recognising numbers 1 to 10	.832	.837
B28	Use one-to-one correspondence	.814	.827
B31	Able to say which number is bigger of two	.810	.823
B21	Write own name in English	.784	.781
B29	Able to count to 20	.784	.798
B19	Aware of writing directions in English	.766	.768
B27	Sort and classify objects	.760	.778
B13	Showing awareness of rhyming words	.749	.776
B32	Recognise geometric shapes	.735	.757
B8	Knows how to handle a book	.734	.746
B14	Participate in group reading activities	.730	.763
B24	Able to remember things easily	.705	.746
B20	Interested in writing voluntarily	.670	.705
B33	Understands simple time concepts	.663	.712
B18	Experimenting with writing tools	.620	.630

Table 4: Factor loadings for the 26-items from the language and cognitive skills domain





#### Communication skills and general knowledge domain

This domain is made up of just 8-items. Although this factor has a similar number of items to the Physical Health and Wellbeing domain, it hangs together better with factor loadings varying from 0.76 to 0.95, suggesting a single cohesive factor. The factor loadings are very consistent across the two waves of data, suggesting the relationship between items and the communication skills and general knowledge construct is stable over time.

Item code	Item description	2009 AvEDI data	2012 AvEDI data
B5	Communicate needs to peers and adults	.946	.949
B1	Use language effectively in English	.935	.930
B3	Ability to tell a story	.929	.935
B6	Understand on first try what is being said	.923	.933
B2	Ability to listen in English	.886	.892
B7	Articulate clearly	.880	.900
B4	Take part in imaginary play	.804	.840
C24	Shows knowledge about the world	.764	.766

Table 5: Factor loadings for the 8-items from the communication skills and general knowledge domain

As mentioned previously, three sets of parameters from the factor analyses are used in the computer simulation. The factor loadings have been described above. The residual variances are inversely related to the factor loadings so they have not been presented. Put simply, items that have high factor loadings have low residual variances suggesting a small proportion of the variance is left over after the variance associated with the latent construct is accounted for. The third set of parameters used in the computer simulation is the threshold values. A brief explanation of these is provided below.

All of the items in the AvEDI are categorical with either 2 or 3 response options. The thresholds provide information about how to convert each of the categorical items to a normally distributed variable, and visaversa. This is important because in the computer simulation, the AvEDI items are first estimated as continuous variables<sup>12</sup> and then re-coded to categorical items. The thresholds indicate where to "cut" the normally distributed variables so that the resulting categorical variables will have the same structure as in the AvEDI data. For example, for item A2 we might observe a NO response for 6.7% of the children and a YES response for 93.3% of the children. The output provided in the threshold section converts these proportions to a z-score. For item A2 the threshold of -1.543, corresponds to a z-score where 93.3% of the normal distribution is to the right of this score. In the computer simulation, a continuous score for item A2 will be generated and then split into a categorical variable by re-coding all scores less than -1.543 to NO and all scores above -1.543 to YES. This step will be discussed in more detail in the following section.

<sup>&</sup>lt;sup>12</sup> A **continuous variable** can take any value in a range of values including decimals. For instance, age is a continuous variable as a child could be 1.83 years of age or 2.45 years old. A continuous age variable could be converted into a **categorical** variable, if we were just interested in age (in whole years). A categorical age variable would take a discrete set of possible options (e.g., 0,1,2,3,4 years of age).





## Stage two: Run the computer simulation

The computer simulations were run in SPSS Statistics Version 19.0 (IBM Corporation, 2011) and were repeated for each wave of data collection (2009 and 2012) and set of specific community sizes. To decide on the community sizes to enter into the simulation, we split the 2009 AEDC national dataset into deciles (groups of ten) based on community size (see Table 6). For instance, 10% of communities across Australia have between 15 and 28 children and 10% of communities have between 1484 and 4545 children. We selected the lower bound for each decile (15, 29, 51, 82 etc.) and ran simulations on each of these 10 community sizes. We supplemented these 10 community sizes with three extra data points at the lower end (n = 20, 26 and 35) to provide additional information. To describe the process of running the computer simulation, we will use the example of the 2009 AEDC data and a community with 35 children.

Communities (deciles)	Number of children
1 (smallest communities)	15–28
2	29-50
3	51-81
4	82-126
5	127-182
6	183-261
7	267-426
8	434-743
9	754-1437
10 (largest communities)	1484-4545

Table 6: Deciles of community size for 2009 AEDC data	

#### Step 1: Generate true scores for the five domains

For each of the five domains, we generate a 'true score'<sup>13</sup> for each of the 35 children by drawing a random score from a standard normal distribution, N(0,1). This set of true scores is replicated 1,000 times so that each child has the same true scores for each of the 1,000 iterations. A transformation is run on these scores to make domain scores correlate with each other as they do in the observed data. The transformed domain scores are then used in the calculation of each of the individual AvEDI items.

<sup>&</sup>lt;sup>13</sup> A **true score** refers to an individual's score on the latent factor (e.g., Physical Health and Wellbeing) assuming that this could be observed/measured directly with no error.





#### Step 2: Generate the observed scores for each of the items

For each of the 96 items in the AvEDI, we generate a continuous score based on the factor loading, error variance estimate and teacher/class level variation and then re-code it to a categorical variable based on the thresholds from the factor analysis. For each of the 1,000 simulations, a child has the same true score on the domain but a different observed score on each item. This simulates the process of 1,000 different teachers assessing the same groups of children on the AvEDI.

The continuous score for each item is computed by adding together three components, which incorporates the variance associated with the teacher/class level variation or measurement error<sup>14</sup>.

- <u>Teacher/class level variation</u>. We draw a random number from a normal distribution with mean = 0, and SD = 0.4, where 0.4 represents a moderate level of teacher level variation. Scores on this component will be centred on zero but each of the 35 children will have a different score to one another and each child will have a different score for the 1,000 replications.
- 2) <u>Teacher/class level variation in consistency of scoring.</u> We draw a random number from a normal distribution with mean = factor loading for that item, and SD = 0.4, where 0.4 represents a moderate level of teacher/class level variation. This random number is multiplied by the 'true score' for the child on the relevant domain from step 1.
- 3) <u>Additional random measurement error.</u> We draw a random number from a normal distribution with mean = 0, and SD = square root of the estimated residual variance for that item from the factor analysis. Scores on this component will be centred on zero but each of the 35 children will have a different score to one another and each child will have a different score for the 1,000 replications.

Scores on these three components are summed together to create a continuous score for each item. The continuous scores are recoded into categorical scores using the thresholds from the factor analysis. For example, in the 2009 AEDC data for item A2, 6.7% of the children were scored YES and 93.3% of the children were scored NO. The threshold for this item was -1.543. In the computer simulation, the continuous score is split into a categorical variable by re-coding all scores less than -1.543 to YES and all scores above -1.543 to NO.

<sup>&</sup>lt;sup>14</sup> A normal distribution has a single peak and is symmetrically distributed with a bell-shaped curve.





#### Step 3: Combine the 96-items into the domain scores and compute all categorical indicators

The 96 items are combined to create the five domain scores according to the standard AvEDI syntax. The Australian cut-offs are applied so that each child is classified as to whether they are developmental vulnerable (1 =Yes, 0 =No), at risk (1 =Yes, 0 =No), and on track (1 =Yes, 0 =No) on each of the five domains. These categorical scores are combined to generate three more indicators, (1) vulnerable on 1 or more domains, (2) vulnerable on 2 or more domains, and (3) On Track of 5 domains.

**Note**. When extending the critical difference methodology in 2016 and 2021, we utilised the simulated data files created in Step 1 and 2 in 2013. We calculated the new AEDC indicators (e.g., On Track on 5 domains), and then proceeded with the remainder of the steps set out in this section. This approach allowed us to explore the critical difference for new AEDC indicators, without making any changes the critical difference estimates for the "vulnerable" indicators that had previously been reported in AEDC products, such as the 2012 and 2015 AEDC National Reports, and the Community Profiles





#### Step 4: Examine the distributions for each of the indicators and extract the SD estimate

The final step is to explore the percentage of the 35 children who are vulnerable, at risk and on track for each of the 5 domains, as well as vulnerable on 1+ domains, 2+ domains and On Track on five domains, across the 1,000 replications. Figure 1 shows the distributions for the percentage of children who are "vulnerable" on the physical health and wellbeing and social competence domains for communities with 35 children. In both cases, the mean level of developmental vulnerability is about 10%. The SD is much larger for the physical health and wellbeing domain (SD = 3.78) than the social competence domain (SD = 2.36). This suggests that we can be less confident about estimating the level of vulnerability in the physical health and wellbeing domain. This is consistent with the results of the factor analysis, which suggested that the social competence domain "hung together" better than the physical health and wellbeing domain.



#### Figure 1: Distribution for % vulnerable on two AvEDI domains for communities with 35 children

Even within a single domain, the simulation will give different SD estimates each time we run it. For example, in Figure 2 the SD for Physical health and wellbeing is 3.91 compared to 3.78 in Figure 1. For this reason, we run the simulation 10 times under each condition (i.e. indicator and community size) and take the average SD from the 10 estimates and use this in Step 3.







#### Figure 2: Distribution for % vulnerable for communities with 35 children (alternate run)





# Stage three: Estimate the critical difference for all community sizes

Step 3 was run in Microsoft Excel. . For each of the 13 community sizes (see Table 6), we ran the computer simulation, examined the distributions for all indicators and extracted the SD from each one. We repeated this process 10 times, and calculated the average standard deviation (SD) for each of the indicators.

Table 7 shows an example of the results for the percentage of children vulnerable on the social competence domain in a community with 35 children. Runs 3 and 10 are shown in Figure 1 and Figure 2, respectively. For Wave 1, on the first run of the simulation the average level of developmental vulnerability for the social competence domain was 6.11% and the SD was 2.32. Over the 10 runs of the computer simulation for Wave 1 (2009), the average level of vulnerability on the Social competence domain was 10.74% and the SD was 2.62. This entire process is repeated for Wave 2 (2012) data, to give an average level of vulnerability of 8.40% and SD of 2.26.

	Wave 1 (2	2009)	Wave 2 (20	)12)
Simulation	Mean	SD	Mean	SD
1	6.11	2.32	7.94	1.70
2	18.57	2.64	2.00	1.44
3	15.47	3.06	12.29	2.71
4	10.19	3.01	7.43	2.28
5	4.69	2.34	3.99	2.71
6	13.99	2.25	10.87	2.50
7	8.39	2.69	9.39	2.51
8	7.78	2.24	8.09	2.02
9	10.35	2.84	4.13	1.60
10	11.83	2.79	17.87	3.13
Average	10.74	2.62	8.40	2.26

Table 7: Computer simulation results – Developmental vulnerability on the social competence domain

The next step was to calculate the standard deviation of the difference in level of developmental vulnerability between the two waves. The formula to calculate the SD of the difference is presented below. Based on the data in Table 7, the SD of the difference is 3.46.

$$SD \ of \ Difference = \sqrt{SD1^2 + SD2^2}$$





Finally, we calculate the critical difference based on the formula for the 95% confidence interval of the difference in the level of developmental vulnerability between the two waves. Based on the data in Table 7, the critical difference is 6.78.

#### Critical difference = 1.96 \* SD of Difference

The null hypothesis is that the 'true difference' between the proportion vulnerable in Wave 1 and 2 is zero. The observed difference in this simulation is 2.34 (10.74-8.40) and there will be some variation around this difference estimate. The critical difference gives us an estimate of this variation. In this example, the critical difference is 6.78. If we observe a difference greater than or equal to the critical difference of 6.78 in a community of 35 children, we can reject the null hypothesis that the true difference is zero and we can conclude that this is a significant difference between the populations at the two time points.

The critical difference was estimated for each of the 13 community sizes, following the same process as describe above. Table 8 shows the critical differences for the percentage of children vulnerable the social competence domain. For communities with 15 children, they need to see a difference in the level of developmental vulnerability of 10 percentage points or more to be statistically significant. For instance, if 20% of children were vulnerable at Wave 1, this would need to drop to below 10% or rise to above 30% at Wave 2 to be significant. However, in larger communities a much smaller change would be significant. In a community of 754 children, a drop from 20% vulnerability at Wave 1 to 18% vulnerability at Wave 2 (or a rise to 22% vulnerable) would be significant as it exceeds the critical difference of 1.4% points.

Community size	
(number of children)	Critical difference
15	10.0
20	8.9
26	6.7
29	7.4
35	6.8
51	5.4
82	4.4
127	3.5
183	2.9
267	2.5
434	2.0
754	1.4
1484	1.0

Table 8: Critical differences – Developmental vulnerability on the Social competence domain





			(	Critical difference			
Community size	Physical health	Social	Emotional	Language and	Communication	Vulnerable on	Vulnerable on
(number of children)	and wellbeing	competence	maturity	cognitive skills	skills and general	1 or more	2 or more
					knowledge	domains	domains
15	14.7	10.0	12.3	11.0	13.7	17.7	12.6
20	12.9	8.9	11.0	9.9	11.2	15.5	10.5
26	11.1	6.7	8.9	8.0	9.8	13.2	9.1
29	10.4	7.4	8.6	7.9	9.6	12.5	9.7
35	9.9	6.8	7.4	6.8	8.7	11.1	8.0
51	8.1	5.4	6.1	5.5	6.9	9.3	6.7
82	6.7	4.4	4.8	5.0	5.6	7.5	5.7
127	5.2	3.5	3.9	4.0	4.4	6.0	4.3
183	4.2	2.9	3.3	3.1	3.7	4.9	3.6
267	3.6	2.5	2.7	2.7	3.2	4.2	3.1
434	2.8	2.0	2.1	2.1	2.4	3.2	2.3
754	2.2	1.4	1.6	1.5	1.9	2.5	1.8
1484	1.5	1.0	1.1	1.1	1.3	1.7	1.3

Table 9: Critical differences for all "vulnerable" based indicators for the AvEDI

Note. This table is used to explain the *process* of calculating the critical difference values but should not be used in practice.

Instead, use the power functions and look up tables presented on pages 36-41.





Table 9 presents the critical difference estimates for all "vulnerable" based indicators. The critical difference estimates were largest for small communities and smallest for larger communities. Given that the number of items in each domain varies between 8 and 26 items, and some domains are more cohesive than others, the critical difference estimates vary between the five different domain indicators. Of the five developmental domains, the critical difference values were smallest in the *social competence* domain. This is not surprising as this domain has a large number of items (24-items) and all items have high loadings (>0.70) on the construct. The *language and cognitive skills* domain generates the second lowest critical values, which is also consistent with the results of the factor analyses where all 26 items have a high loading on the factor (> .62). The critical difference values are a little higher for the *emotional maturity* domain. This may be related to the small number of items that measure Anxious and Fearful behaviour that have lower factor loadings on the construct (<.60). The developmental domains with the higher critical difference values are *physical health and wellbeing* and *communication skills and general knowledge*. The larger critical difference values are related to the smaller number of items measuring these domains, and lower factor loadings for some items in the *physical health and wellbeing* domain

The critical difference estimates for "vulnerable on 1 or more domains" are larger than the critical difference estimates for any of the indicators for the individual domains. This composite indicator incorporates measurement error from all of the five domains, and thus the critical difference needed to conclude a significant change is larger. The indicator "vulnerable on 2 or more domains" also incorporates measurement error from all 5 domains but generates lower critical difference estimates than those for "vulnerable on 1 or more domains". Given that the domains are correlated, once we know that a child is vulnerable on 1 domain, this increases the likelihood that they are vulnerable on other domains. This inter-relatedness of the domains is one of the reasons that the critical differences for "vulnerable on 2 or more domains" are smaller than the critical differences for "vulnerable on 1 or more".

While Table 9 is useful, and a necessary first step, it is static and does not allow us to calculate an exact critical difference for each different size community in Australia. If a community has 600 children, what would the critical difference value be for the Social competence domain? To answer this question, we plotted the community size against the critical difference and examined the curve. The resulting curve for the social competence domain is shown in Figure 3.









Figure 3: Plot of critical differences x community size – % vulnerable in the Social competence

Figure 3 allows us to select the number of children in the community and approximate the critical difference. For instance, if we draw a vertical line up from 600 until it hits the curve and then a horizontal line across to the y-axis, we can see that the critical difference for a community with 600 children is about 1.6. In Stage 4, we describe a method of quantifying this curve so that we can more accurately estimate the critical difference value for any community size in Australia.







# Stage four: Estimate the power function for each of the indicators

Step 4 was run in Microsoft Excel. From Table 9, we plotted the critical difference against the community size for each of the indicators and examined the curves. Following the method used at the University of British Columbia, we fitted a power function to the curve and estimated the equation of this curve.

The power function takes the form shown below, where Y is the critical difference and X is the number of children in the community. K and A will change for each of the indicators.

$$Y = K * (x)^{-A}$$

K can be interpreted as the critical difference for communities with 1 child. It does not make practical sense to use the K value in this way because we would never report AEDC results for a community with one child. However, at a glance, we can compare the K values for the different indicators to see which generate higher or lower critical difference values (higher K = higher critical values).

As an example, from Figure 4 below, K = 56.20, and A = 0.493. If a community had 1,500 children, then X = 1,500, and the critical difference could be calculated from the equation below.

$$Y = 56.20 * (1,500)^{-0.493} = 1.53$$

Over the next few pages, we present the curves and equations for each of the indicators.







#### Power functions: % vulnerable



#### Figure 4: Critical difference for % vulnerable on the Physical Health and Wellbeing domain

Figure 5: Critical difference for % vulnerable on the Social Competence domain

**Discover, Prevent, Cure,** 











Figure 7: Critical difference for % vulnerable on the Language and Cognitive Skills domain







Figure 8: Critical difference for % vulnerable on the Communication and General Knowledge domain









#### Power functions: % at risk



Figure 9: Critical difference for % at risk on the Physical Health and Wellbeing domain



Figure 10: Critical difference for % at risk on the Social Competence domain























Figure 13: Critical difference for % at risk on the Communication and General Knowledge domain







#### Power functions: % on track



Figure 14: Critical difference for % on track on the Physical Health and Wellbeing domain



Figure 15: Critical difference for % on track on the Social Competence domain









Figure 16: Critical difference for % on track on the Emotional Maturity domain













Figure 18: Critical difference for % on track on the Communication and General Knowledge domain







#### Power functions: Summary indicators















Figure 21: Critical difference for On Track on 5 domains (OT5)



#### Power functions: All indicators

	Category of interest				
Developmental Domain	Vulnerable	At Risk	On Track		
Physical Health and Wellbeing	$Y = 56.200 * (x)^{-0.493}$	$Y = 69.543 * (x)^{-0.495}$	$Y = 65.824 * (x)^{-0.494}$		
Social Competence	$Y = 37.135 * (x)^{-0.487}$	$Y = 62.313 * (x)^{-0.491}$	$Y = 48.218 * (x)^{-0.486}$		
Emotional Maturity	$Y = 48.062 * (x)^{-0.515}$	$Y = 69.417 * (x)^{-0.501}$	$Y = 50.303 * (x)^{-0.486}$		
Language and Cognitive skills	$Y = 41.927 * (x)^{-0.497}$	$Y = 67.780 * (x)^{-0.498}$	$Y = 51.978 * (x)^{-0.488}$		
Communication and General Knowledge	$Y = 50.933 * (x)^{-0.500}$	$Y = 73.207 * (x)^{-0.499}$	$Y = 59.974 * (x)^{-0.497}$		

Table 10: Power functions for each of the 5 developmental domains for each of the 3 categories

Table 10 presents the power functions for each of the five developmental domains and each of three categories of interest. In general, the critical difference values are a little higher for the on track category than the vulnerable category, and are highest overall for the at risk category. This suggests that there is more uncertainly around the estimate of the percentage of children who are" at risk" than the percentage of children who are "on track" or "vulnerable". The power functions for the two summary indicators (vulnerable on 1 or more domains, and vulnerable on 2 or more domains) are presented in Table 11.

Table 11: Power functions to calculate the critical difference for the 2 summary indicators

Summary indicators	Power functions
Vulnerable on 1 or more domains (DV1)	$Y = 67.888 * (x)^{-0.502}$
Vulnerable on 2 or more domains (DV2)	$Y = 47.639 * (x)^{-0.495}$
On Track on 5 domains (OT5)	$Y = 63.635 * (x)^{-0.491}$





# Using the critical difference in practice

#### Using the power functions

The power functions allow the calculation of a critical difference for any specific community size and indicator. For example, the critical difference for the % of children vulnerable on the Social Competence domain for a community with 600 children is 1.65 (see below).

 $Y = 37.135 * (600)^{-0.487} = 1.65$ 

Most communities will have a different number of children in different years (2009, 2012, and 2015). If a community had 600 kids in 2009 and 650 in 2015, the <u>smaller</u> number should be used into the critical difference formula (i.e. 600 kids).

These power functions are used to calculate the critical difference for communities that are presented on the <u>AEDC Data Explorer<sup>15</sup></u>. The power functions can be copied into an excel spreadsheet and used to calculate the critical difference for schools or communities based on their size. This and will provide a more precise estimate of the critical difference than using a look up table (see below).

#### Using the Comparative Results Tool

An excel tool has been developed to calculate the critical difference for any of the 18 AEDC indicators and for a community/group of any size. The tool can be accessed on the AEDC website (<u>www.aedc.gov.au</u>). Data users need to:

- 1. Select the AEDC indicator of interest (e.g., On Track on 5 domains)
- 2. Select the two AEDC collection cycles they want to compare (e.g., 2009 and 2021)
- 3. Enter in the % of children in the relevant category (e.g., On track on 5 domains) at each time point
- 4. Enter the number of children in the community at each time point

Based on this information, the tool will calculate the critical difference and the observed difference in scores on the AEDC indicator between the two waves of data, and provide some interpretive text.



<sup>&</sup>lt;sup>15</sup> <u>https://www.aedc.gov.au/data/data-explorer</u>







#### Figure 22. Comparative Results Tool (screen shot)

#### Using look up tables

While the power functions and Comparative Results Tool are useful for data managers and people with access to excel, they may not be accessible to all data users. As such, a set of lookup tables with critical difference values for communities of different sizes will be provided in the 2015 AEDC Community Profiles. In each case, the critical difference is based on the lower bound of the community size. That is, for a community with 200-299 children, we have calculated the critical difference for a community with 200 children. The look up tables are presented below.

$$Y = 37.135 * (200)^{-0.487} = 2.81$$





#### Table 12. Look up table of critical differences for five AvEDI domains - On track indicators

	Developmentally on track critical difference percentage points				
Community size	Dhysical boolth	Social compotence	Emotional maturity		Communication and
(number of children)	and wellbeing (%)	(%)	(%)	skills (school-based) (%)	general knowledge (%)
15-19	17.3	12.9	13.5	13.9	15.6
20-24	15.0	11.2	11.7	12.0	13.5
25-29	13.4	10.1	10.5	10.8	12.1
30-39	12.3	9.2	9.6	9.9	11.1
40-59	10.6	8.0	8.4	8.6	9.6
60-79	8.7	6.6	6.9	7.0	7.8
80-99	7.6	5.7	6.0	6.1	6.8
100-199	6.8	5.1	5.4	5.5	6.1
200-299	4.8	3.7	3.8	3.9	4.3
300-699	3.9	3.0	3.1	3.2	3.5
700-1499	2.6	2.0	2.1	2.1	2.3
1500-2499	1.8	1.4	1.4	1.5	1.6
2500-3499	1.4	1.1	1.1	1.1	1.2
3500-9999	1.2	0.9	1.0	1.0	1.0
10,000-100,000	0.7	0.5	0.6	0.6	0.6
100,000-200,000	0.2	0.2	0.2	0.2	0.2
200,000-300,000	0.2	0.1	0.1	0.1	0.1
300,000+	0.1	0.1	0.1	0.1	0.1





	Developmentally at risk critical difference percentage points				
<b>Community size</b> (number of children)	Physical Health and wellbeing (%)	Social competence (%)	Emotional maturity (%)	Language and cognitive skills (%)	Communication and general knowledge (%)
15-19	18.2	16.5	17.9	17.6	19.0
20-24	15.8	14.3	15.5	15.2	16.4
25-29	14.1	12.8	13.8	13.6	14.7
30-39	12.9	11.7	12.6	12.5	13.4
40-59	11.2	10.2	10.9	10.8	11.6
60-79	9.2	8.3	8.9	8.8	9.5
80-99	7.9	7.2	7.7	7.6	8.2
100-199	7.1	6.5	6.9	6.8	7.4
200-299	5.0	4.6	4.9	4.8	5.2
300-699	4.1	3.8	4.0	4.0	4.3
700-1499	2.7	2.5	2.6	2.6	2.8
1500-2499	1.9	1.7	1.8	1.8	1.9
2500-3499	1.4	1.3	1.4	1.4	1.5
3500-9999	1.2	1.1	1.2	1.2	1.2
10,000-100,000	0.7	0.7	0.7	0.7	0.7
100,000-200,000	0.2	0.2	0.2	0.2	0.2
200,000-300,000	0.2	0.2	0.2	0.2	0.2
300,000+	0.1	0.1	0.1	0.1	0.1

#### Table 13. Look up table of critical differences for five AvEDI domains – At risk indicators





	Developmentally vulnerable critical difference percentage points					
Community size	Physical health	Social competence	Emotional maturity	Language and	Communication and	
(number of children)	and wendering (%)	(%)	(%)	cognitive skills (%)	general knowledge (%)	
15-19	14.8	9.9	11.9	10.9	13.2	
20-24	12.8	8.6	10.3	9.5	11.4	
25-29	11.5	7.7	9.2	8.5	10.2	
30-39	10.5	7.1	8.3	7.7	9.3	
40-59	9.1	6.2	7.2	6.7	8.1	
60-79	7.5	5.1	5.8	5.5	6.6	
80-99	6.5	4.4	5.0	4.7	5.7	
100-199	5.8	3.9	4.5	4.3	5.1	
200-299	4.1	2.8	3.1	3.0	3.6	
300-699	3.4	2.3	2.5	2.5	2.9	
700-1499	2.2	1.5	1.6	1.6	1.9	
1500-2499	1.5	1.1	1.1	1.1	1.3	
2500-3499	1.2	0.8	0.9	0.9	1.0	
3500-9999	1.0	0.7	0.7	0.7	0.9	
10,000-100,000	0.6	0.4	0.4	0.4	0.5	
100,000-200,000	0.2	0.1	0.1	0.1	0.2	
200,000-300,000	0.1	0.1	0.1	0.1	0.1	
300,000+	0.1	0.1	0.1	0.1	0.1	

# Table 14. Look up table of critical differences for five AvEDI domains – Vulnerable indicators





	Summary indictors - Critical difference percentage points					
<b>Community size</b> (number of children)	Vulnerable on 1 or more domains (%)	Vulnerable on 2 or more domains (%)	On Track on 5 domains (%)			
15-19	17.4	12.5	16.8			
20-24	15.1	10.8	14.6			
25-29	13.5	9.7	13.1			
30-39	12.3	8.8	12.0			
40-59	10.7	7.7	10.4			
60-79	8.7	6.3	8.5			
80-99	7.5	5.4	7.4			
100-199	6.7	4.9	6.6			
200-299	4.7	3.5	4.7			
300-699	3.9	2.8	3.9			
700-1499	2.5	1.9	2.6			
1500-2499	1.7	1.3	1.8			
2500-3499	1.3	1.0	1.4			
3500-9999	1.1	0.8	1.2			
10,000-100,000	0.7	0.5	0.7			
100,000-200,000	0.2	0.2	0.2			
200,000-300,000	0.1	0.1	0.2			
300,000+	0.1	0.1	0.1			

Table 15. Look up table of critical differences for summary indicators





# **Summary of findings**

- The critical difference methodology was developed at the University of British Columbia and has been adapted for use with the AEDC data
- The methodology involves exploring the psychometric properties of the AvEDI and running a computer simulation to mimic the process of 1,000 teachers completing the AvEDI for the same group of children.
- The critical difference indicates the minimum difference in AEDC indictors between two waves of AEDC data needed to conclude a significant difference.
- The critical difference depends on the number of children in a community and the level of measurement error in the AEDC indicator.
- The five domains of the AvEDI have different levels of measurement error associated with them and therefore generate different critical difference estimates.
- The relationship between the critical difference and community size can be described by a power function.
- There is a different power function for each of the AEDC indicators
- The power function provides the capacity to calculate the critical difference for any community size based on the number of children in that community.







## References

- Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizbility: A liberalization of reliability theory. *British Journal of Statistical Psychology*, *16*, 137-163.
- Forer, B., Guhn, M., & Zumbo, B. (2011). Using a Generalizability Theory Framework to Investigate the Reliability of Aggregated Early Development Instrument Change Scores. Paper presented at the Society for Research in Child Development Biennial Meeting, Montreal, Quebec, Canada.

IBM Corporation. (2011). IBM SPSS for Windows. Version 19.0. Armonk, NY: IBM.

Muthen, L. K., & Muthen, B. O. (1998-2006). Mplus User's Guide. Fourth Edition. Los Angeles, CA: Muthen & Muthen.



