Using measures of quality to improve the learning outcomes of all children

Dr Dan Cloney
Australian Council for Educational Research

Dr Dan Cloney is a research fellow in the Centre for Education Policy and Practice at the Australian Council for Educational Research. Dan's expertise is in early education, cognitive development, academic achievement and inequality.

Dan's most recent work includes longitudinal studies in Australia and Indonesia. One of these, the Overcoming Disadvantage in Early Childhood Project, is a partnership with the Australian Literacy & Numeracy Foundation to study the effectiveness of a preschool intervention in New South Wales. Another, the UNICEF Modelling of Universal Pre-Primary Education, is a study of the impact of investment in early childhood programs in the Regency of Bogor, Indonesia. Dan was also a research fellow on the E4Kids study, which is Australia's largest longitudinal study of the contribution of early childhood programs to children's learning and development.

Dan's recent publications focus on the potential for high-quality early childhood programs to reduce SES-related achievement gaps.

Abstract

There is compelling evidence that high-quality early childhood education and care (ECEC) programs can act to narrow achievement gaps attributed to social inequality. This evidence is typically observed in model programs, designed by experts and offered to vulnerable families outside the market. In everyday settings, where market forces may price families out of certain programs or poor local availability may preclude attendance, ECEC programs do not appear to deliver these significant gains or close these gaps. There is a need to continually improve quality in all ECEC settings to deliver on the potential of early education. It is unclear, however, how quality improvement can be achieved in way that will deliver the best start in life for all.

This paper looks at what early childhood interaction quality looks like right now in Australian services and internationally. Specifically, what educator practices are related to children's learning and development? This paper contributes to the knowledge about quality improvement in two ways. The first is related to how educators can leverage research into pedagogical quality to collect data and improve their own practice. The second relates to the organisation of the ECEC system and how it must be arranged to ensure all children get equal access to high-quality ECEC experiences. Together, these two contributions have the potential to increase the effect ECEC programs have on children's learning and development outcomes and to deliver on the promise of narrowing achievement gaps and breaking the link.

This paper explores the challenge of lifting the quality of Australian ECEC programs, so that the system can deliver on the promise of reducing achievement gaps related to disadvantage. Specifically, this paper brings together the latest research to ask how large-scale psychometric analysis can be used at the classroom level for educators, or communities of educators, to: (1) collect their own data about their practice, (2) visualise it on a continuum of pedagogical quality, and (3) use this information to demonstrate growth in quality.

Introduction

There is evidence of the potential for high-quality early childhood education and care (ECEC) programs to reduce and even close achievement gaps attributed to relative disadvantage. This is a key part of efforts to reduce unjust, unnecessary and preventable inequities caused by entrenched and intergenerational socio-demographic circumstances (Goldfeld et al., 2017, 2018). Participation in model ECEC programs – specifically designed by experts and provided to vulnerable families outside the everyday market – is associated with significant and life-long benefits (Schweinhart, 2005). In carefully designed studies, greater developmental gains are seen for children in high-quality ECEC programs when compared to low-quality programs (Burchinal et al., 2008; Duncan & Sojourner, 2013). These studies, however do not address the fact that the everyday market tends to produce lower quality programs than seen in the model programs; ECEC programs in the US, UK and Australia demonstrate that some aspects of ECEC quality are low across the entire population (Tayler, Ishimine, Cloney, Cleveland, & Thorpe, 2013), and that children from disadvantaged backgrounds face significant barriers to accessing high-quality programs or any programs at all (Cloney, Cleveland, Hattie, & Tayler, 2016; Cloney, Cleveland, Tayler, Hattie, & Adams, 2017a; Hatfield, Lower, Cassidy, & Faldowski, 2015).

The implication of these patterns is that ECEC programs do not appear to be delivering on their potential to reduce inequality. In Australia, for example, children's early oral language skills vary significantly at age three and those who are behind early, continue to be behind (or potentially even further behind) when they enter school (Tayler, Cloney, & Niklas, 2015). These early gaps are strongly associated with later gaps in school achievement. For example, children who had low, compared to average, oral language skills at 3 years of age scored significantly lower on the National Assessment Program – Literacy and Numeracy (NAPLAN) domains of reading, writing and language conventions at Grade 3 (approximately five years later). The magnitude of this difference is as large: 90 scale points (spelling), which is more than 1.5 standard deviations (Tayler et al., 2016b).

This paper presents a way of addressing these issues, and ensuring that ECEC programs live up to their potential to close achievement gaps related to inequity or disadvantage. A method is presented to use data from large-scale research to produce a practitioner-focused quality improvement tool. By allowing educators to locate their current practice on an empirically validated continuum of instructional quality, they can undertake appropriately targeted, incremental quality improvement. Improvements in instructional quality are known to contribute to children's learning and development outcomes.

Measuring early childhood education and care quality – what is important?

This paper focuses on the quality of instruction and brings together the findings from new research (in particular, Cloney & Hollingsworth, manuscript submitted for publication), to answer two questions: first, can we produce a reliable and valid continuum of instructional quality using the Classroom Assessment Scoring System (CLASS) (Pianta, La Paro, & Hamre, 2008), and second, how can this continuum be used to lift the quality of instruction in Australian ECEC programs?

The analysis in this paper uses data from a large, longitudinal study designed to estimate what early childhood programs add to children's learning and development from the ages of three to eight years in Australia. The CLASS is a tool that measures the quality of teacher–student interactions in the classroom. Measures of interaction quality are widely used by researchers and by governments and are shown to be predictive of children's learning and development (Sabol, Soliday Hong, Pianta, & Burchinal, 2013). The CLASS yields ratings on a scale of one to seven on three domains: emotional support, classroom organisation and instructional support. This paper only focuses on instructional support. This domain is theoretically and empirically associated with children's cognitive and academic achievement in early childhood and school literature (Hamre et al., 2013; Pianta & Hamre, 2009; Pianta et al., 2008).

The positive relationship between instructional support and early outcomes is despite the observed pattern of instructional support being scored low in the population of ECEC programs. Observed effect sizes (ES) in everyday settings for reading and language outcomes range from very small (figure significant but effect size not stated in Hamre et al., 2013) to small (ES = 0.23) (Burchinal, Vandergrift, Pianta, & Mashburn, 2010). Observed ES in everyday settings for mathematics outcomes are larger (up to ES = 0.34) (Burchinal et al., 2010). A US study found that instructional support only predicted reading and mathematics achievement above a threshold of 3.25 (on a seven-point scale) (Burchinal et al., 2010) while the average score observed in the study was only 2.05 (1.4 SD lower than the threshold) (Tayler et al., 2013). In studies where the threshold was set lower (e.g. to ensure sufficient numbers in the contrast group) effects were not as clear. A threshold of 3 (out of 7) has shown no effect for early literacy and a small effect for inhibitory control (ES = 0.23) (Hatfield et al., 2016). There is a clear opportunity to push instructional quality in all settings up to levels known to have impacts on learning and development.

Method

Participants

More than 2500 children participated in this study, and the study protocol provides specific details of the sampling and measurement used (Tayler et al., 2016a).

The data presented here are from the 2011 wave of data collection and include observations of 993 classrooms including: International Standard Classification of Education (ISCED) (UNESCO, 2012) level 010 (e.g. child care), ISCED 020 (pre-primary programs for mostly four year olds), and the first year of ISCED 100 (e.g., the first year of school for mostly 5 year olds) that were located in 647 individual schools or services). Home-based services were excluded.

Instrumentation

The CLASS measures three domains: emotional support, Classroom Organization, and instructional support. The observer scores 10 indicators nested within these three domains (four, three, and three indicators within each domain, respectively). To get a score for each domain, an observer scores each indicator on a scale where 1–2 is low quality; 3–5 is mid quality; and 6–7 is high quality. The domains and indicators that make up CLASS are described in Table 1 (p. 86). Note that these are the descriptions for the pre-K version of the measure (generally for children aged three to five); however, this analysis also includes ratings on the K–3 measure that includes the same domains and indicators; however the description of the indicators is changed to be contextually appropriate to the age group (Pianta et al., 2008).

This paper only focuses the instructional support domain, and the estimated scale reliability is 0.89 (Cloney et al., 2015a, 2017). Additional information regarding the training of observers, inter-rater reliability, and model estimation can be found in Cloney et al. (2017b).

Analytical approach

The data from the 993 classrooms is modelled as a multidimensional partial credit model (Adams, Wilson, & Wang, 1997). All three CLASS domains are modelled simultaneously, but only the instructional support continuum is presented here. This continuum is different from the instructional support score given from the CLASS manual as it takes in to account measurement error as well estimating the relative difficulty of each of the indicators in order to place classrooms on the continuum.

Table 1 Description of the indicators (dimensions) and factors (domains) of the instructional support domain of the Classroom Assessment Scoring System

Domain	Dimension	Description
Instructional support	Concept development	Measures the teacher's use of instructional discussions and activities to promote students' higher-order thinking skills and cognition and the teacher's focus on understanding rather than on rote instruction.
	Quality of feedback	Assesses the degree to which the teacher provides feedback that expands learning and understanding and encourages continued participation.
	Language Modeling	Captures the quality and amount of the teacher's use of language-stimulation and language-facilitation techniques.

Table adapted from Pianta et al., 2008

Results

Instructional support continuum

This study builds on the results published in Cloney and Hollingsworth (manuscript submitted for publication) and Cloney et al. (2015a, 2015b, 2017b), which illustrate how more advanced modelling techniques could overcome the psychometric limitations of CLASS already reported in the literature (e.g., Pakarinen et al., 2010). The key finding of these papers is that the underlying theorisation of the CLASS is sound, and the orthodox measurement approach is limited.

Figure 1 (p. 87) is an item map that summarises the findings of the model for instructional support. The vertical dotted line represents the continuum of instructional support from low (bottom of the figure) to high (top of the figure). The dots represent the observed distributions of classrooms on this continuum. Three columns, for the indicators that make up instructional support, show the thresholds for the indicators. For example, CD.2 means the second threshold for concept development. This is the location on the instructional support continuum at which a classroom or an educator would have a more than 50 per cent chance of scoring 3 or higher on concept development (note that the first threshold, CD.1 represents the location where a classroom or educator would have a greater than 50 per cent chance of being scored 2 or higher: there are 7 – 1 = 6 thresholds in the model). For this paper, it is important to note that that the threshold CD.2 represents a classroom moving from low- to mid-quality on concept development and yet the distribution of observed classrooms (representative of classrooms in

Australia) shows approximately three-quarters of classrooms operate below this level (the 75th centile is -0.59 logits). This implies that concept development is difficult to exhibit – rarely observed – in classrooms. Only classrooms scoring very high on the instructional support continuum could be expected to be demonstrating the behaviours described in the upper indictors of concept development. Compare it with progressing from low- to mid-quality for, Language Modeling (LM.2: more than 50 per cent of classrooms operate above this threshold on the continuum of instructional quality).

From these results, it is clear that the behaviours described in the CLASS manual are not equally difficult to demonstrate at each scoring point – that is, scoring a 1 on CD is not the same as scoring a 1 on LM, rather, some behaviours are more challenging to demonstrate that others. This is shown in Table 2. Of note, CD appears to be the most difficult to demonstrate (e.g. the relative size of the low scores, shaded in grey) of the indicators.

Conclusion

This study demonstrates that a continuum of instructional quality can be produced from the CLASS measure. A major barrier, however, to the use of the CLASS measure is that it requires significant training to yield data and analysis to produce interpretable results – such a process is more aligned with research or with monitoring than it is for use in the classroom (Cloney & Hollingsworth, manuscript submitted for publication). This discussion, therefore, considers how the results presented above can be applied in the quality improvement efforts of educators.
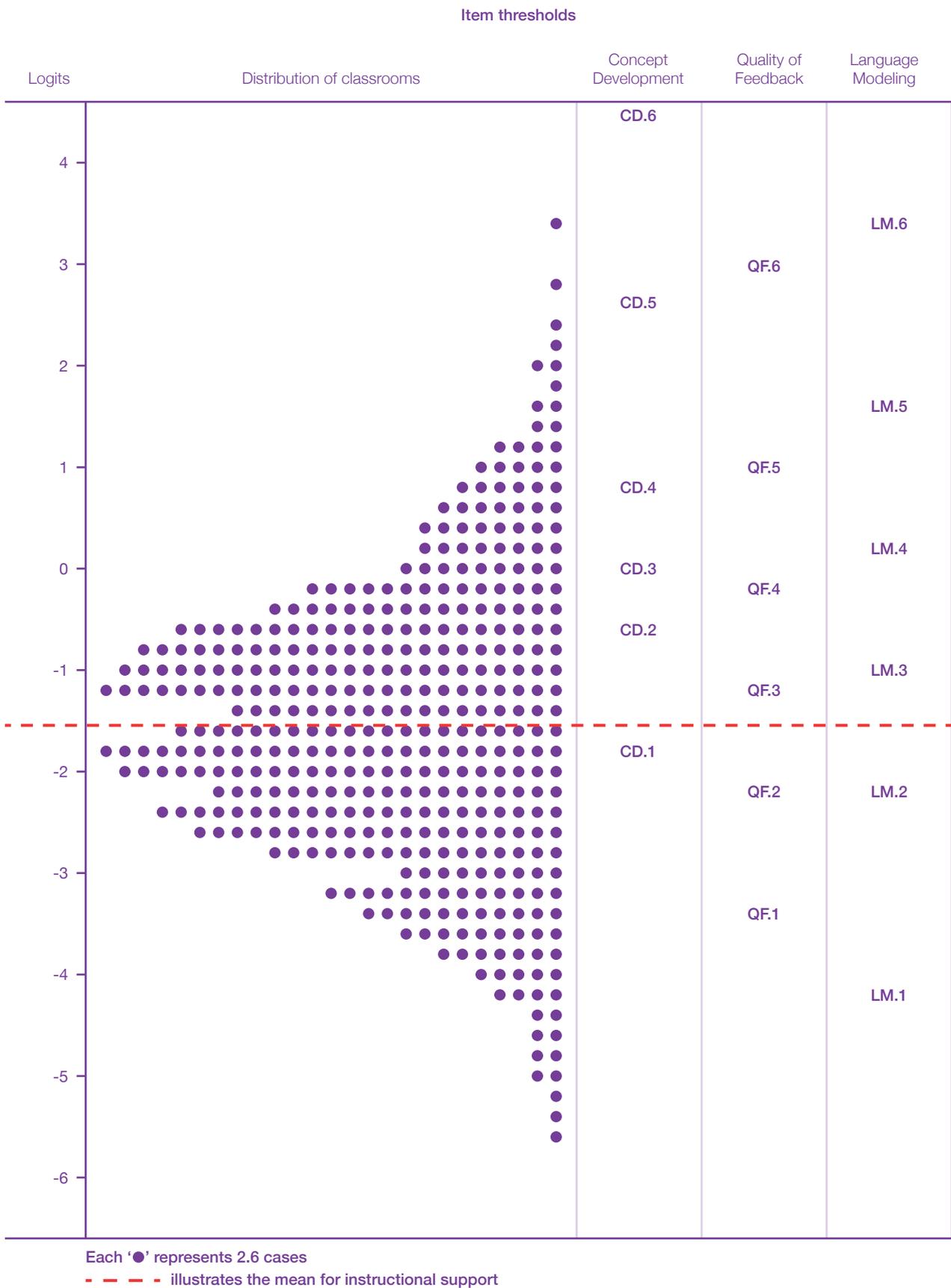
Figure 1 Wright map of multidimensional partial credit model of CLASS instructional support

Table 2 Illustration of the continuum of behaviours making up instructional support using evidence from IRT analysis

Logits	Concept development[1]	Quality of feedback[2]	Language modeling[3]
4	The educator often provides opportunities for students to be creative and/or generate their own ideas and products exemplified by brainstorming, planning and producing. The educator consistently links concepts and activities to one another and to previous learning.	The educator often provides additional information to expand the students understanding or actions exemplified by expansion, clarification and specific feedback. The educator often offers encouragement of the student's efforts that increases student's involvement and persistence exemplified by recognition, reinforcement and observed student persistence.	There are frequent conversations in the classroom exemplified by back-and-forth exchanges, contingent responding and peer conversations. The educator often repeats or extends the students' responses including elaboration.
3			
2			
1	The educator sometimes provides opportunities for students to be creative and/or generate their own ideas and products exemplified by brainstorming, planning and producing. The educator sometimes links concepts and activities to one another and to previous learning.	The educator occasionally provides additional information to expand the students understanding or actions exemplified by expansion, clarification and specific feedback. The educator occasionally offers encouragement of the student's efforts that increases student's involvement and persistence exemplified by recognition, reinforcement and observed student persistence.	There are limited conversations in the classroom. The educator sometimes repeats ort extends the students' responses.
0			
-1			
-2	The educator rarely provides opportunities for students to be creative and/or generate their own ideas and products. Concepts and activities and presented independent of one another, and student are not asked to apply previous learning.	The educator rarely provides additional information to expand the students understanding or actions. The educator rarely offers encouragement of students' efforts that increases students' involvement and persistence.	There are few, if any, conversations in the classroom. The educator rarely, if ever, repeats or extends the students' responses.
-3			
-4			
-5			
-6			

Descriptions are adapted from Pianta et al., 2008.

---

[1] Descriptions of the behavioural markers of *creating*, and *integration* are given. The other behavioural markers within this dimension are *analysis and reasoning, and connection to the real world*.

[2] Descriptions of the behavioural markers of *providing information* and, *encouragement and affirmation* are given. The other behavioural markers within this dimension are *scaffolding*, feedback *loops*, and *prompting thought processes*.

[3] Descriptions of the behavioural markers of *frequent conversation*, and *repetition and extension* are given. The other behavioural markers within this dimension are *self- and parallel talk*, and *open-ended questions*, and *advanced language*.

The instructional support continuum can be used to focus and support the efforts of educators. It provides a clear understanding of what quality looks like, can be used to locate the level of practice, and can be used in everyday settings. The continuum provides educators with a map showing levels of practice from low to high quality. By qualitatively comparing their own practice to the described continuum, educators can locate themselves and the set of behaviours proximal (but above) their own level of practice. This is the target area for quality improvement and can be done without the need for complex analysis.

An applied example of this can be seen through the lens of the Structured Stimulation of Teacher Reflection (SSTR) approach (Hollingsworth & Clarke, 2017). The instructional support continuum is the element of practice to plan for. Educators should then proceeds plan around demonstrating this focus and record their practice (e.g., through peer observation or recording). Following this, the educator can use the recording or peer observation to collect evidence of their behaviours as described on the continuum ('What I saw', My thoughts about what I saw', and 'What I might do differently'), and then engage in a professional conversation about their practice.

The innovative part about this process of quality improvement is that the cycle of reflexive practice is centred on strong empirical evidence – the educator seeks to locate themselves on the continuum and improve their practice by moving from their current location to the next highest location on continuum. Educators can see, therefore, that if they are located at the lowest levels on the instructional support continuum than demonstrating higher levels of concept development is, on average very difficult. Educators will likely be demonstrating more growth in Language Modeling than in concept development at this location on the continuum. This approach means that unrealistic quality improvement or professional learning targets are not set and that growth is structured in a way that is commonly observed in other similar classrooms.

Most importantly, this focus on quality improvement is most likely to be associated with growth in children's learning and development. Ensuring that quality improvement efforts focus on aspects of quality that are most likely to yield improved growth for children is critical for the ECEC market to deliver on its potential. This is most important for programs providing services to low income or vulnerable children. At present, we know these programs are the lowest quality (on average) and should be provided with the most support to lift quality to levels likely to narrow persistent achievement gaps. This is the only way to ensure that all children have the best start in life and to eliminate inequities that are unjust, unnecessary and preventable, and caused by entrenched sociodemographic circumstances.

References

Adams, R. J., Wilson, M. R., and Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, *21*, 1–23.

Burchinal, M., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher–child interactions and instruction. *Applied Developmental Science*, 12(3), 140–153. doi:10.1080/10888690802199418

Burchinal, M., Vandergrift, N., Pianta, R. C., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, *25*(2), 166-176. doi: 10.1016/J.Ecresq.2009.10.004

Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychological Review*, *101*(4), 568-586. doi: 10.1037/0033-295x.101.4.568

Bruner, J. S. (1978). The role of dialogue in language acquisition. In A. Sinclair, R., J. Jarvelle, & W.J.M. Levelt (Eds.), *The child's concept of language*. New York, NY: Springer-Verlag.

Cloney, D., Cleveland, G., Hattie, J., & Tayler, C. (2016). Variations in the availability and quality of early childhood education and care by socioeconomic status of neighbourhoods. *Early Education and Development*, *27*(3), 384-401. doi:10.1080/10409289.2015.1076674

Cloney, D., Cleveland, G., Tayler, C., Hattie, J., & Adams, R. (2017a). The selection of ECEC programs by Australian families: Quality, availability, usage and family demographics. *Australasian Journal of Early Childhood Research*, *41*(4).

Cloney, D., & Hollingsworth, H. (Manuscript submitted for publication). Using measures of pedagogical quality to provide feedback and improve practice. In J. Lodge, J. Cooney Horvath, & L. Corrin (Eds.), *From data and analytics to the classroom: Translating learning analytics for teachers*. Routledge.

Cloney, D., Nguyen, C., Adams, R., Tayler, C., Cleveland, G., & Thorpe, K. (2017b). Psychometric properties of the Classroom Assessment Scoring System (Pre-K): Implications for measuring interaction quality in diverse early childhood settings. *Journal of Applied Measurement*, *18*(3).

Cloney, D., Nguyen, C., Adams, R., Tayler, C., Cleveland, G., & Thorpe, K. (2015a, 14 Jan). *Measuring early childhood education and care classroom quality: Psychometric properties of classrooms assessment scoring system*. Paper presented at the Sixth International Conference on Probabilistic Models for Measurement in Education, Psychology, Social Science and Health, Cape Town.

Cloney, D., Nguyen, C., Adams, R., Tayler, C., & Cleveland, G. (2015b, 17 Apr). *The psychometric properties of the Classroom Assessment Scoring System (CLASS)*. Paper presented at the American Educational Research Association (AERA) 2015 annual meeting: Toward justice: Culture, language, and heritage in education research and praxis. Chicago, IL.

Duncan, G. J., & Sojourner, A. J. (2013). Can intensive early childhood intervention programs eliminate income-based cognitive and achievement gaps? *Journal of Human Resources*, *48*(4), 945-968. doi:10.3368/jhr.48.4.945

Goldfeld, S., O'Connor, M., Cloney, D., Gray, S., Redmond, G., Badland, H., Kochanoff, A. T. (2017). Understanding child disadvantage from a social determinants perspective. *Journal of Epidemiology and Community Health*. doi: 10.1136/jech-2017-209036

Goldfeld, S., O'Connor, M., O'Connor, E., Chong, S., Badland, H., Woolfenden, S., Mensah, F. (2018). More than a snapshot in time: Pathways of disadvantage over childhood. *International Journal of Epidemiology*, doi: 10.1093/ije/dyy086

Hamre, B. K., Pianta, R. C., Jason, T. D., DeCoster, J., Mashburn, A. J., Jones, S. M., Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal*, *113*(4), 461–487. doi:10.1086/669616

Hatfield, B. E., Burchinal, M. R., Pianta, R. C., & Sideris, J. (2016). Thresholds in the association between quality of teacher–child interactions and preschool children's school readiness skills. *Early Childhood Research Quarterly*, *36*, 561–571. doi: 10.1016/j.ecresq.2015.09.005

Hatfield, B. E., Lower, J. K., Cassidy, D. J., & Faldowski, R. A. (2015). Inequities in access to quality early care and education: Associations with funding and community context. *Early Childhood Research Quarterly*, *30*, Part B, 316–326. doi:10.1016/j.ecresq.2014.01.001

Hollingsworth, H., & Clarke, D.J. (2017). Video as a tool for focusing teacher self-reflection: Supporting and provoking teacher learning. *Journal of Mathematics Teacher Education*. doi: 10.1007/s10857-017-9380-4.

Pakarinen, E., Lerkkanen, M. K., Poikkeus, A. M., Kiuru, N., Siekkinen, M., Rasku-Puttonen, H., & Nurmi, J. E. (2010). A validation of the classroom assessment scoring system in Finnish kindergartens. *Early Education and Development*, *21*(1), 95–124. doi:10.1080/10409280902858764

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, *38*(2), 109–119. doi: 10.3102/0013189x09332374

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system manual: Preschool (Pre-K) version*. Baltimore, MD: Brookes.

Schweinhart, L. J. (2005). *Lifetime effects: the High/Scope Perry preschool study through age 40*. Ypsilanti, Mich.: High/Scope Press.

Sabol, T. J., Soliday Hong, S. L., Pianta, R. C., & Burchinal, M. R. (2013). Education. Can rating pre-K programs predict children's learning? *Science* (New York, N.Y.), *341*(6148), 845-846. doi: 10.1126/science.1233517

Tayler, C., Cloney, D., Adams, R., Ishimine, K., Thorpe, K., & Nguyen, C. (2016a). Assessing the effectiveness of Australian early childhood education and care experiences: Study protocol. *BMC Public Health*, *16*(1), 1–12. doi: 10.1186/s12889-016-2985-1

Tayler, C., Cloney, D., & Niklas, F. (2015). A bird in the hand: Understanding the trajectories of development of young children and the need for action to improve outcomes. *Australasian Journal of Early Childhood*, *40*(3), 51–60.

Tayler, C., Cloney, D., Niklas, F., Cohrssen, C., Thorpe, K., Page, J., & D'Aprano, A. (2016b). *Final report to the partner organisations for the Effective Early Education Experiences (E4Kids) study*. Melbourne Graduate School of Education. doi: 10.4225/49/58f99f47a2ab4

Tayler, C., Ishimine, K., Cloney, D., Cleveland, G., & Thorpe, K. (2013). The quality of early childhood education and care services in Australia. *Australasian Journal of Early Childhood*, *38*(2), 13–21.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2012). *International Standard Classification of Education ISCED 2011*. Montreal, Canada: UNESCO Institute for Statistics. Retrieved from http://uis.unesco.org/en/topic/international-standard-classification-education-isced