



Comparing regulatory and non-regulatory indices of early childhood education and care (ECEC) quality in the Australian early childhood sector

Iram Siraj^{1,2} · Steven J. Howard¹ · Denise Kingston³ · Cathrine Neilsen-Hewett¹ · Edward C. Melhuish^{1,2,4} · Marc de Rosnay¹

Received: 2 October 2018 / Accepted: 3 April 2019 / Published online: 16 April 2019
© The Author(s) 2019

Abstract

This study examines associations between Australia’s regulatory ratings of quality in early childhood education and care (ECEC)—the National Quality Standard (NQS)—and two research-based quality rating scales. The analytic sample consisted of 257 ECEC services across three Australian states. Results indicated (1) modest positive associations between NQS ratings and scale scores; (2) some specificity between NQS quality areas (educational programs and practice; relationships with children) and one research scale—the Sustained Shared Thinking and Emotional Wellbeing (SSTEW) scale; (3) variability in quality scales scores within each NQS designation; and (4) mitigation of these associations when the time-gap between ratings exceeded 24 months. Findings suggest NQS and research scales tap some common core of quality, yet capture different aspects of quality, suggesting both could be used to raise standards of quality in Australian preschools, where the research scales potentiate raising quality to even higher levels than NQS.

Keywords National Quality Standard · ECERS-E · SSTEW · Quality · PRESCHOOL · Regulation

✉ Iram Siraj
iram.siraj@education.ox.ac.uk

¹ Early Start, Faculty of Social Sciences, University of Wollongong, Wollongong, NSW 2522, Australia

² Department of Education, University of Oxford, Oxford OX2 6PY, UK

³ Department of Education, University of Sussex, Brighton BN1 9RH, UK

⁴ Department of Psychological Sciences, Birkbeck, University of London, London WC1E 7HX, UK

Introduction

Evidence supports the importance of high-quality educational experiences in early childhood education and care (ECEC) settings (e.g. OECD 2012). Its benefits include the potential to promote children's short- and long-term development, such as educational success (Melhuish et al. 2015), the possibility of reducing inequalities linked to socio-economic status (Siraj and Mayo 2014), and the potential to foster economic growth in the longer term (Ho et al. 2010).

Despite such benefits being well recognised (Melhuish et al. 2015; Pianta 2012; Sylva et al. 2014; Taylor et al. 2016), there continues to be debate around how high-quality experiences can be ensured and maximised. Defining ECEC pedagogies and practices that demonstrably support and enhance children's learning and development remains a research imperative. However, the term quality itself is a contested concept, due mostly to its subjective nature. Numerous studies have shown that stakeholders (e.g. parents, educators, children and governments) hold different views of quality and value different characteristics of practice (Education Review Office 2010; Penn 1996). For example, parents value the proximity of an ECEC setting to their homes and perceived happiness of children (Plantenga 2011), while many governments (including the Australian government) emphasise impact on children's outcomes (their social, emotional and cognitive development). A comprehensive discussion of the multiple perspectives of quality can be found elsewhere (Kingston 2017; Kingston and Melvin 2012; Mathers et al. 2012; Moss and Dahlberg 2008).

While it is noted that quality is a complex term, the current study considers the Australian government's national measure of quality—the National Quality Standard (NQS) assessment—which, amongst other important roles (informing policy, ensuring public accountability and supporting children's and families' entitlements to high-quality ECEC provision) measures and monitors quality with a focus on the impact ECEC provision can have on children's outcomes (ACECQA 2017a). Given the lack of information on the extent to which NQS results align with those shown to predict child outcomes, we compared the NQS with two quality rating scales [The Sustained Shared thinking and Emotional Well-being (SSTEW) and Early Childhood Environment Rating Scale—Extension (ECERS-E) (Siraj et al. 2015; Sylva et al. 2003)]. Drawing on data from our studies of quality in ECEC and government NQS ratings, we considered how NQS ratings compared these two quality rating scales with known associations with children's outcomes (social, emotional and cognitive development) (Howard et al. 2018; Sylva et al. 2006).

International efforts to capture and improve ECEC quality

Governments around the world try to ensure that their processes and measures to monitor and improve ECEC services deliver high-quality environments for children and families. For instance, the UK government has scrutinised quality by revamping its inspection systems and curricula for children below age 7 (Parker 2013). New Zealand, having achieved a robust revision for the Te Whariki curriculum with

educational outcomes for children and near-universal provision for pre-schoolers (NZ Ministry of Education 2017), government attention has now turned to quality.

Similarly, in 2009, Australian state governments established a National Quality Framework (NQF) for all ECEC. This took effect in 2012 and was designed to support greater consistency across state and territory educational systems, supported also by implementing a national Early Years Learning Framework (EYLF; DEEWR 2009) and a National Quality Standard (NQS) assessment and rating system (ACECQA 2017a). Australia's NQS was developed by a group of early childhood experts, in consultation with government, to establish and maintain the quality of ECEC in Australia via a comprehensive regulatory framework (see Table 1).

In common with government inspection and monitoring processes internationally, NQS is based on judgments of trained field staff (assessors) who evaluate a broad range of elements to assign ratings which conceptually capture overall service quality—including quality associated with learning and quality associated with compliance. Services evaluated include home-based family day care, long-day care (who care for infants, toddlers and preschool-aged children) and dedicated preschool programs (focused on the year before school).

NQS has focussed ECEC educators and providers on service quality. Yet regulatory authorities have understandably placed an early emphasis on child safety and regulatory compliance, as is captured in a number of NQS quality areas. However, NQS quality ratings are often interpreted in the field as interchangeable with *research* measurements of quality ECEC, which have been shown empirically to correspond with improved child outcomes in learning and development (Howard et al. 2018; O'Connell et al. 2016). Although NQS was informed by research (OECD 2006) and is linked to the EYLF, there remains limited publicly available evidence to evaluate whether it can differentiate between settings that differ in their impact on children's learning and development. The capacity to make this differentiation is particularly significant for children from disadvantaged backgrounds, for whom there is compelling evidence that higher quality ECEC can enhance development in a myriad of domains (Melhuish et al. 2015).

Conceptualising quality associated with improved child outcomes in ECEC

Recent reviews (Melhuish et al. 2015; Siraj and Kingston 2015) and studies (Taylor et al. 2016) have identified characteristics important for enhancing children's learning, development and wellbeing through ECEC, namely.

1. Adult–child interactions which are sensitive, warm and emotionally supportive;
2. well-trained staff, including teachers and directors with relevant degree and post-graduate qualifications;
3. a developmentally appropriate curriculum and educationally orientated focus, including support for learning within play and play-based activities;
4. ratios and group sizes which allow staff to interact with children regularly and deeply;
5. leadership which supports collaboration and maintains consistency in care quality;

Table 1 Areas of quality covered by the NQS, SSTEW and ECERS-E

NQS 7: Quality areas (QA)	SSTEW subscales	ECERS-E subscales
QA1. Aspects of educational programs and practices for children	Subscale 1: building trust, confidence and independence	Subscale 1: literacy
QA2. Health and safety	Subscale 2: social and emotional wellbeing	Subscale 2: mathematics
QA3. Aspects of the physical environment	Subscale 3: supporting and extending language and communication	Subscale 3: science and environment
QA4. Staffing	Subscale 4: supporting learning and critical thinking	Subscale 4: diversity
QA5. relationships with children	Subscale 5: assessing learning and language	
QA6. Partnerships with families and communities		
QA7. The way services are managed and led		

6. staff development which ensures continuity, stability and improving quality;
7. facilities which support health and safety, and are accessible to parents;
8. sharing educational goals with parents/carers, and supporting the home learning environment.

These characteristics show the complexity of promoting stronger cognitive and social–emotional outcomes for children. Indeed, there appears to be an inter-play between aspects of *structural quality* known to be associated with high-quality ECEC (e.g. group size, child/adult ratios, educator qualifications) (Howes et al. 2008; Slot et al. 2015) and aspects of *process quality* concerned with a child’s everyday lived experiences (e.g. the activities, opportunities and interactions available within a setting, the educators and the other children, and the available and accessible materials) (Howes et al. 2008). This distinction between *structural* and *process* quality helps to identify the aspects of overall quality which most influence children’s early learning and development (Donabedian 1980; Siraj et al. 2017b).

As *structural* quality can be measured more objectively and easily, its aspects are most frequently researched and regulated, and now dominate government ECEC inspection and monitoring around the world. The manner in which different aspects of structural quality exert influence is not simple, and many recent studies have suggested an overreliance on structural notions of quality may be misplaced (e.g. Slot et al. 2015). Increasingly, international research shows that the *process* of adult–child and child–child interactions predicts children’s outcomes most powerfully (Melhuish et al. 2015; Siraj and Kingston 2015).

In summarising available research on process quality, the OECD (2012) reported that educators’ sensitivity and responsiveness, the quality of their interactions, and their ability to extend and scaffold children’s learning and thinking were critical to children’s outcomes. These aspects of process quality, while more difficult to capture than structural aspects, have been the focus of sustained empirical investigation using quality rating scales in ECEC contexts. It is, therefore, important to ask how quality defined by regulatory frameworks corresponds with quality measured on such scales.

Quality measured by quality rating scales

High-quality interactions, defined as those that support and extend children’s thinking, have been increasingly linked to child outcomes i.e. their social–emotional and cognitive development (Pianta 2012; Sylva et al. 2014). These interactions, termed *sustained shared thinking* (SST) by Siraj-Blatchford et al. (2015), are conceptualised as interactions in which “...two or more individuals ‘work together’ in an intellectual way to solve a problem, clarify a concept, evaluate an activity, extend a narrative, and so forth. Both parties must contribute to the thinking, and it must develop and extend the understanding” (Siraj-Blatchford et al. 2002, p. 8). While SST’s influence is reflected in the Australian EYLF and other curricula internationally (e.g. English Early Years Foundation Stage 2012), practices associated with SST remain poorly understood and observed infrequently (Sylva et al. 2004; Taylor

et al. 2016). It should thus be a priority to support the understanding and practice of SST, and measure and monitor it accurately.

These newer understandings about the important adult role in fostering interactions have prompted the development of tools designed to specify and capture pedagogies and practices associated with process aspects of quality. Accordingly, the SSTEW scale (Siraj et al. 2015) was designed to support, increase and improve the identification and practice of high-quality interactions within ECEC settings. Similarly, the ECERS-E aims to capture aspects of curricular quality, such as mathematical, scientific and literacy learning, and diversity. These instruments have an international reputation for (i) measuring important aspects of ECEC quality which relate to key domains of development (e.g. language, numeracy, science, diversity, self-regulation, social–emotional wellbeing) and what constitutes effective practice in each; (ii) the standardisation processes they have undergone; and (iii) their well-established psychometric properties (e.g. predictive validity of child outcomes such as language and numeracy) (Howard et al. 2018; Mathers et al. 2012; Sylva et al. 2004). Quality rating scales are widely used in studies across Europe, Asia Pacific regions and beyond, and have also been used to capture quality, and changes in quality, following professional development in Australia (e.g. Siraj et al. 2018).

Reconciling NQS and quality rating scales

In contrast, the NQS is a regulatory tool covering seven quality areas and 18 standards (2–3 in each quality area), which broadly consider important aspects of structural and process quality alongside child safety and regulatory compliance. The statutory NQS assessment and rating of ECEC services in Australia, conducted by government-authorised assessors, result in centres receiving individual ratings for each quality area, and an overall quality rating. There are five possible ratings: significant improvement required, working towards NQS, meets NQS, exceeds NQS and excellent. The lowest and highest ratings are rarely given.

There appear to be important areas of overlap and difference between NQS and the scales. In terms of similarity, elements of NQS quality area 1 (educational program and practice) show the greatest synergy with quality scale indicators which emphasise curricula process quality. For example, NQS quality area 1 requires judgments about curricula and their implementation, the educators' abilities to take a child-centred approach, and promotion of children's agency with a foundation in children's interests, ideas, knowledge and culture. This includes intentional teaching, planning and assessment at an individual level, support for all children to participate and fostering caregivers' awareness about the planning for, and the progress of, their children. There are also similarities between SSTEW and NQS quality area 5 (relationships with children). Both stress building strong and respectful relationships to support collaborative learning and child independence; and both recognise the importance of partnership with parents and the early home learning environment (Melhuish et al. 2008).

Yet, NQS and quality scales differ in the scope and depth of quality covered. ECERS-E, for example, includes maths, science and literacy, whereas NQS

uses these terms infrequently (only three mentions of each) and positions them as optional (e.g. potential examples of practice). In summary, while the quality rating scales focus most highly on aspects of process quality, NQS conceptualises quality much more broadly, with a focus on structural and regulatory compliance (health and safety, educators' qualification levels, adult/child ratios, etc.). As such, while the scales have an evidence base linking better ratings to better child outcomes, the extent to which resultant NQS ratings identify centres that differ in their children's outcomes remains unclear.

The current study

This study thus begins a process of empirical evaluation that should accompany any large-scale regulatory system (indeed similar processes have been initiated in the UK and US; Mathers et al. 2012; Sabol et al. 2013). Specifically, this study compared NQS with two quality rating scales: ECERS-E and SSTEW, combining data gathered during large-scale projects in Australia and from published regulatory NQS ratings. These particular scales were selected because they focus on curricular quality (corresponding to NQS quality area 1) and interactional quality (corresponding to NQS quality area 5). Associations were examined between the quality rating scales scores and the overall and specific-quality-area NQS ratings. Consideration was also given to potential effects of service location (i.e. state), the time since NQS assessment (i.e. within 2 years of quality scale observations), and the variability in quality within NQS designations. It was expected that NQS and quality scale ratings would be distinct (i.e. capturing different aspects of quality) yet related (i.e. sharing a common core of quality), given at least some overlap between the indices, while intentionally different in breadth and aims.

Method

Sample

Analyses were conducted on NQS and quality rating scale data from 257 ECEC services across three Australian states. Centres were selected for representation, not representativeness, across a range of centre characteristics. Largely consistent with national distributions, there were more centres in metropolitan ($n = 156$, 60.7%) than non-metropolitan regions ($n = 101$, 29.3%) and more long-day care centres ($n = 221$, 86.0%) than preschools ($n = 36$, 14.0%). The average socio-economic decile of the ECEC services' catchment areas was 3.92 ($SD = 2.28$) (per the Australian Bureau of Statistics' area-level Socio-economic Index for Areas, or SEIFA; ABS 2008), which indicates a slightly lower socio-economic catchment area than would be expected in the broader population. The average number of places per centre was 65.37 ($SD = 28.30$). For overall NQS rating categories, most centres achieved *meeting* ($n = 128$, 49.8%) or *exceeding* designations ($n = 101$, 39.3%). Fewer had *working*

toward ratings ($n=28$, 10.9%), which is consistent with the profile of ratings nationally (ACECQA 2018).

As quality scale observations are conducted on individual rooms, this sample yielded 323 rooms (64 centres had two rooms and one had three rooms). However, each centre received only one overall NQS rating. To circumvent this issue of non-independence of observations in centres with multiple rooms, one room was randomly selected for inclusion from centres with multiple room ratings, yielding 257 independent ratings for the quality scales.

Measures

Details of the main areas of quality covered by the NQS, ECERS-E and SSTEWS can be found on Table 1.

National Quality Standard (NQS) ratings

Australia's NQS assessment and rating is undertaken with all ECEC services across Australia. Fully trained, authorised officers visit each centre to assess against 18 standards in the aforementioned seven quality areas (see Table 1). These individual quality area ratings are then combined into one overall NQS rating for each centre. Ratings are published on national registers, from which the current data were drawn, and displayed in services. Data on centre ratings for this study were collected from the national register in March 2017.

Quality rating scales

ECERS-E (Sylva et al. 2003) measures the quality of curricula, environments and pedagogy in ECEC settings. It comprises 15 items across four subscales. SSTEWS (Siraj et al. 2015) considers practice which supports children to develop skills in sustained shared thinking and emotional wellbeing. It contains 14 items across five subscales. Subscale foci are detailed in Table 1.

ECERS-E and SSTEWS were scored using on-balance judgements derived from a full-day room observation. Each item was scored from 1 (*inadequate* quality) to 7 (*excellent* quality) based on patterns of the presence or absence of each item's indicators. A score of 3 indicates *basic/minimal* quality and 5 indicates *good* quality. Both scales have shown good reliability and predictive validity of children's attainment at school entry (Howard et al. 2018; Sylva et al. 2006). Items in the subscales were averaged to generate subscale scores, and subscales were averaged to yield overall quality scores (Table 2).

The scale assessments were conducted by highly trained observers throughout a 1-day room observation in participating centres. Observers were trained intensively for 5 days, including in-field practice ratings with a highly experienced trainer/observer, followed by inter-rater reliability checks that compared independent ratings from a full-day joint observation with a highly experienced trainer/observer. Observers had to meet the following rigorous inter-rater

Table 2 Mean quality scores (and SDs) for ECERS-E, SSTEW and their respective subscales by state

	Overall	State 1 n = 147	State 2 n = 26	State 3 n = 84
Mean ECERS-E quality score	2.86 (.93)	2.93 (.93)	2.58 (.90)	2.80 (.93)
1. Literacy	3.51 (1.05)	3.62 (1.05)	3.15 (1.19)	3.45 (.98)
2. Mathematics	2.70 (1.24)	2.68 (1.13)	2.54 (1.24)	2.79 (1.43)
3. Science and environment	2.72 (1.16)	2.84 (1.19)	2.46 (1.05)	2.58 (1.12)
4. Diversity	2.49 (1.00)	2.60 (1.07)	2.17 (1.00)	2.39 (.90)
Mean SSTEW quality score	3.66 (1.24)	3.77 (1.17)	3.78 (1.40)	3.43 (1.30)
1. Trust, confidence and independence	4.53 (1.39)	4.62 (1.32)	4.65 (1.61)	4.34 (1.44)
2. Social-emotional wellbeing	4.00 (1.66)	4.07 (1.57)	4.19 (1.70)	3.82 (1.80)
3. Language and communication	4.24 (1.36)	4.35 (1.21)	4.50 (1.53)	3.96 (1.51)
4. Learning and critical thinking	2.71 (1.33)	2.78 (1.34)	2.72 (1.37)	2.58 (1.30)
5. Assessing learning and language	2.82 (1.40)	3.01 (1.45)	2.84 (1.52)	2.46 (1.20)

ECERS-E and SSTEW are measured on a 7-point ECEC quality scale, 1 = inadequate, 3 = basic/minimal, 5 = good, 7 = excellent

reliability standard prior to data collection in field: (1) an intra-class correlation exceeding .70 ($M = .86$); (2) a correlation exceeding .70 ($M = .86$); (3) a mean difference in scores less than .75 ($M = .43$); and (4) a score agreement (within 1 point) of at least 80% ($M = 93\%$).

Centre characteristics

Information on centres' geographic region, service type, SEIFA decile and service size were collected at the time of scale observations. From national registers, we also recorded the time elapsed since NQS rating (max = 4.05 years prior to our observations).

Results

Initial data exploration

While it was not expected that the quality ratings scales would differ by state, preliminary analyses sought to establish this before exclusion of this factor from subsequent analysis (Table 2). Regression analyses showed that the addition of state (coded as dummy variables) to a model of SEIFA, geographic region, service type and maximum number of places predicting ECERS-E scores did not improve model fit, $\Delta F(2, 249) = .29$, $p = .752$. This was also the case for SSTEW, $\Delta F(2, 249) = .58$, $p = .564$. As there were no systematic effects of state on ECERS-E or SSTEW scores, state was omitted from further analyses.

Differences in quality rating scale scores between NQS rating categories

ECERS-E

To analyse potential associations between ECERS-E and NQS ratings, we ran hierarchical multiple regressions to investigate the extent to which NQS ratings (quality area 1 and overall) predicted ECERS-E scores, controlling for SEIFA, geographic region, service type and number of places (see Table 3). NQS was recoded into dummy variables, with *meeting* as the reference category. The model at the first step, including only control variables, was significant, $F(5, 251) = 3.11$, $p = .010$, although no predictor made a significant independent contribution to ECERS-E quality score. This was identical when considering overall NQS rating or NQS rating for quality area 1.

Of interest for the current investigation, the second step included NQS ratings to investigate whether ECERS-E scores improved with improved NQS rating. This full model was significant for overall NQS ratings, $F(7, 249) = 4.16$, $p < .001$, and for NQS quality area 1 ratings, $F(7, 249) = 4.00$, $p < .001$. The addition of NQS rating at step 2 improved the model (see Table 4), and confirmed that centres receiving *exceeding* on NQS overall (or on quality area 1) achieved significantly higher ECERS-E scores than those receiving *meeting*. There were no statistically significant differences in ECERS-E scores between the centres receiving *working toward* and *meeting* NQS ratings. These findings thus indicated an association between NQS designations and quality rating scores, yet this association was broad (not specific to quality area 1) and only for higher quality NQS designations. Indeed, identical regression analyses for the other NQS quality area ratings (quality area 2 through quality area 7) showed similar results, with centres

Table 3 Mean ECERS-E and ECERS-E Subscale Quality Scores (and SDs) for NQS ratings (working toward, meeting, exceeding) in quality area 1 and overall rating

	NQS rating		
	Working toward	Meeting	Exceeding
Quality area 1: educational program and practice ^a			
Mean ECERS-E quality score	2.45 (.60)	2.74 (.89)	3.16 (.99)
1. Literacy	3.03 (.78)	3.39 (1.01)	3.86 (1.09)
2. Mathematics	2.27 (.69)	2.65 (1.29)	2.90 (1.24)
3. Science and environment	2.27 (.85)	2.56 (1.04)	3.11 (1.33)
4. Diversity	2.22 (.85)	2.37 (.92)	2.77 (1.11)
Overall NQS rating			
Mean ECERS-E quality score	2.61 (.72)	2.69 (.83)	3.14 (1.02)
1. Literacy	3.16 (.76)	3.32 (.98)	3.85 (1.12)
2. Mathematics	2.56 (1.19)	2.55 (1.16)	2.93 (1.34)
3. Science and environment	2.37 (.81)	2.53 (1.02)	3.05 (1.32)
4. Diversity	2.33 (.90)	2.34 (.95)	2.72 (1.06)

^aCentres for NQS Quality Area 1 categories are 20 working toward, 153 meeting and 84 exceeding

Table 4 Hierarchical multiple regression results for ECERS-E Quality Scores regressed on NQS (quality area 1 and overall) for full sample (N=257) and the reduced (within 24 months) sample (N=184)

	National quality standard							
	Quality area 1: educational program and practice				Overall NQS rating			
	N=257		N=184		N=257		N=184	
	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β
	.04**		.07**		.05**		.09**	
SEIFA		.08		.07		.08		.07
Geographic region								
Metro versus inner regional		.09		.08		.09		.08
Metro versus outer regional		-.08		-.07		-.08		-.08
Service type (LDC vs. pre)		.10		.12		.11		.15
Maximum places		-.08		-.08		-.06		-.06
NQS rating								
Meeting versus working toward		-.09		-.08		-.03		-.06
Meeting versus exceeding		.17**		.24**		.21**		.28**

* $p < .05$; ** $p < .01$

rated *exceeding* receiving significantly higher ECERS-E scores for every analysis (β s ranged from .17 to .21).

However, the mean ECERS-E scores for *exceeding* centres were at the *basic/minimal* level (a score of 3 out of a possible 7) according to ECERS-E ratings (see Table 3). There was also substantial variability within NQS levels (Table 4): for *working toward*, ECERS-E scores ranged from 1.67 to 4.29 ($M=2.61$, $SD=.72$); for *meeting*, ECERS-E scores ranged from 1.29 to 4.63 ($M=2.69$, $SD=.83$); and for *exceeding*, ECERS-E scores ranged from 1.29 to 6.71 ($M=3.14$, $SD=1.02$).

SSTEW

The same analyses were run to investigate associations between NQS ratings (in quality area 1, quality area 5 and Overall) and SSTEW scores (Table 6). The model at first step was again significant (and identical) for NQS quality area 1, quality area 5 and overall ratings, $F(5, 251)=2.58$, $p=.027$. Only geographic region was significant amongst control factors, such that inner-regional settings had higher SSTEW scores than did centres in metropolitan settings. While this is noteworthy, it was not examined further as these centres were not recruited in a geographically representative manner, and this was not an a priori aim of this study.

The model was significantly improved with the addition of NQS quality area 1 rating, $F(7, 249)=2.86$, $p=.007$, and quality area 5 rating, $F(7, 249)=3.02$, $p=.005$, but not for overall NQS ratings. Inspection of these results indicated that, for NQS quality area 1, centres receiving *exceeding* on NQS achieved significantly higher scores on SSTEW than those receiving *meeting*. There was no

significant difference in SSTEWE ratings between centres receiving *meeting* and *working toward* designations. The same pattern was found for NQS quality area 5. This showed that associations of NQS with SSTEWE were specific to the anticipated NQS quality areas, yet only between higher quality designations. Repeating these analyses for the other quality areas indicated that only NQS quality area 7 was additionally associated with SSTEWE scores, such that centres rated as *exceeding* received significantly higher SSTEWE scores than those rated *meeting* ($\beta = .18, p = .006$). SSTEWE scores were thus related to quality area 1 and quality area 5, as predicted, and were also related to quality area 7.

Again, the mean SSTEWE scores for *exceeding* centres were still below *good* levels (i.e. a score of below 4) according to the SSTEWE scale (see Table 5). There was again substantial variability within NQS levels (Table 5): for *working toward*, SSTEWE scores ranged from 1.48 to 5.98 ($M = 3.41, SD = 1.07$); for *meeting*, SSTEWE scores ranged from 1.05 to 5.88 ($M = 3.51, SD = 1.28$); and for *exceeding*, SSTEWE scores ranged from 1.17 to 6.70 ($M = 3.92, SD = 1.20$).

Table 5 Mean SSTEWE and SSTEWE Subscale Quality Scores (and SDs) for NQS ratings (working toward, meeting, exceeding) in quality area 1, quality area 5 and overall ratings

	NQS rating		
	Working toward	Meeting	Exceeding
Quality area 1: educational program and practice			
Mean SSTEWE quality score	3.27 (1.07)	3.53 (1.25)	4.00 (1.21)
1. Trust, confidence & independence	4.18 (1.34)	4.40 (1.44)	4.86 (1.39)
2. Social-emotional wellbeing	3.45 (1.79)	3.86 (1.67)	4.40 (1.54)
3. Language and communication	3.70 (1.26)	4.15 (1.42)	4.54 (1.19)
4. Learning and critical thinking	2.63 (1.40)	2.53 (1.26)	3.07 (1.39)
5. Assessing learning and language	2.40 (1.11)	2.71 (1.32)	3.10 (1.55)
Quality area 5: relationships with children			
Mean SSTEWE quality score	3.21 (.75)	3.46 (1.25)	3.94 (1.21)
1. Trust, confidence and independence	4.30 (1.07)	4.28 (1.40)	4.86 (1.34)
2. Social-emotional wellbeing	3.54 (1.57)	3.77 (1.74)	4.33 (1.52)
3. Language and communication	3.50 (1.01)	4.04 (1.42)	4.55 (1.24)
4. Learning and critical thinking	2.31 (.70)	2.56 (1.31)	2.94 (1.37)
5. Assessing learning and language	2.36 (.90)	2.67 (1.27)	3.03 (1.56)
Overall NQS rating			
Mean SSTEWE quality rating score	3.41 (1.07)	3.51 (1.28)	3.92 (1.20)
1. Trust, confidence and independence	4.26 (1.07)	4.37 (1.49)	4.81 (1.29)
2. Social-emotional wellbeing	3.75 (1.78)	3.83 (1.69)	4.30 (1.55)
3. Language and communication	3.88 (1.23)	4.14 (1.43)	4.47 (1.26)
4. Learning and critical thinking	2.67 (1.29)	2.51 (1.26)	2.98 (1.40)
5. Assessing learning and language	2.48 (1.11)	2.69 (1.38)	3.06 (1.46)

* $p < .05$; ** $p < .01$

Additional analyses

Given the prevalence of staff turnover, ongoing professional learning and other change factors in the ECEC sector, subsequent analyses sought to evaluate whether the associations between NQS and quality rating scales change when quality ratings were within a reasonable time of each other (i.e. within 2 years). Two years was identified because the time between NQS ratings can be lengthy, yet centres continue to claim their ratings (exceeding, good etc.) even following large-scale changes. These analyses revealed that centres rated less than 24 months before quality ratings ($n=184$ rooms) conformed to the above pattern of findings, except that centres rated more than 24 months prior ($n=73$) showed little association. Subsequent analyses were conducted on this reduced sample, <24 months, and appear in Tables 4 and 6 for ECERS-E and SSTEW. These results suggest that associations between NQS and quality rating scores were strengthened if they occurred within 24 months of each other.

Discussion

The current study sought to evaluate the degree of association between Australia's system of NQS assessment and rating with quality scores from research-based quality rating scales. Results suggested that NQS does indeed capture common elements of quality as these rating scales, yet this association predominantly distinguished between high (exceeding) and lower-quality centres (working toward, meeting), and more highly when NQS ratings had occurred within the past 2 years. Despite this association, quality levels of even *exceeding* services were at basic levels of quality on average, as defined by the quality rating scales. This suggests that NQS may function as an important mechanism to draw attention to quality, and ensure a minimum threshold of quality across the sector, while the scales provide possible tools and direction for centres ready to further extend on this base level of quality. These results echo similar international studies which compare government-authorised monitoring and quality processes with research measures of quality. In England, for example, comparisons between government inspection processes and environmental quality ratings were also modest (Mathers et al. 2012), with similar results found in the US (Sabol et al. 2013).

This study also indicated that there were high levels of variability on scale measures within NQS rating designations, even when considering only centres rated in the previous 24 months. That is, several centres achieving high-quality scores on ECERS-E and SSTEW were rated as not yet meeting the NQS, and the reciprocal pattern was also common. Although the reasons for this are likely multiple and complex, three explanations are proposed. First, there are differences in the quality areas considered. That is, while NQS is necessarily broad in its focus, there is evidence that combining instructional, process and compliance, without highlighting the specific elements linked to child outcomes, can obscure the meaning of ratings generated. As Sabol et al. (2013) reported during a review of monitoring systems in the US, some quality indicators (e.g. adult/

Table 6 Hierarchical multiple regression results for SSTEWE Quality Scores regressed on NQS (quality area 1, 5 and overall) for full sample (N=257) and the reduced (within 24 months) sample (N = 184)

	National Quality Standard											
	Quality area 1: educational program and practice			Quality area 5: relationships with children			Overall NQS rating					
	N=257	N=184		N=257	N=184		N=257	N=184		N=257	N=184	
	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β
SEIFA	.03*	-.02	.03**	-.01	.03*	-.02	.02	-.02	.05***	-.02	.02	-.02
Geographic region												
Metro versus inner regional		.14*		.13		.15*		.14		.14*		.13
Metro versus outer regional		-.03		.04		-.03		.03		-.03		.04
Service type (LDC vs. pre)		.06		.13		.07		.14		.07		.15
Maximum places		-.06		-.03		-.04		-.02		-.05		-.02
NQS rating												
Meeting versus working toward		-.07		-.09		-.04		-.06		-.02		-.10
Meeting versus exceeding		.14*		.19*		.16*		.17*		.13		.19*

* $p < .05$; ** $p < .01$

child interactions) are related to children's learning, but these associations are mitigated when multiple indicators were added to the quality rating, including structural aspects typically found within the government monitoring and inspectional systems under study. In such cases, assigning high-quality designations does not necessarily denote optimal provision for child development and learning. Rather, Sabol et al. (2013) suggest that, where information about multiple aspects of quality is collected, care should be given to separate their analysis and reporting. Further, given the costs associated with collecting multiple quality indicators, there may be economies for focusing separately on indicators that have established associations with learning and child development outcomes. For instance, in the current results, three NQS quality areas related specifically to dimensions of interactional quality as assessed by the SSTEW scale (although no such differentiation across quality areas was evident for ECERS-E). Separate reporting of these process quality dimensions may draw additional focus to important distinctions between aspects of ECEC quality.

Second, while a common core of quality appears to be captured by both types of measures, their associated training, materials and guidance (including how indicators may be understood and interpreted, and how complex concepts such as SST are described) influence the fidelity of their application. Ambiguous guidelines, such as "educators and co-ordinators...promoting a sense of community in the service" (NQS quality area 5; ACECQA 2017b, p. 24), can be interpreted in multiple ways (e.g. embracing cultural diversity, staff togetherness, parent relationships), which could be interpreted as being satisfied by various types, frequencies and quality of practices. This ambiguity introduces inter-rater reliability concerns, and unclear targets for those educators tasked with ensuring that guidelines are met. The scales provide an alternative model in this regard, with indicators and descriptors designed to be concrete and tangible—they can be seen, heard, read, and require specific, observed, well-defined evidence. Although they require some professional judgments, training on principles underlying environmental and pedagogy rating scales and practice in their use ensures inter-rater reliability (Siraj et al. 2017a). Descriptions of the behaviours to be evidenced within the quality rating scales provide clear guidelines, descriptions and indicators of inadequate, minimal, good or excellent practice, with the indicators building upon each other. From this perspective, non-shared variance between NQS and the scales may be, at least in part, due to differing interpretations between assessors or within assessors as their experiences, reference points and interpretations evolve.

Third, there are by design some fundamental differences in the underlying principles and practices considered by these measures of quality. For instance, whereas ECERS-E has two subscales dedicated to mathematics and science and the adult's supportive role in fostering these through adult-guided and play-based experiences, NQS has little direct reference to either area. This may be related to widespread belief in the sector, and reinforced by the NQS, about how children learn best. Indeed, the research underpinning these scales reinforces the importance of a child-centred approach, where educators follow the child's lead and interests during and through play. However, the scales specifically emphasise the importance of leveraging these interests and play experiences to support learning and development

through SST, and in areas such as emergent mathematics, science and exploration. At present, there is sparse mention of these practices and domains in the NQS.

Also complicating associations between these measures is the high rate of turnover of staff in the ECEC sector. Together, these factors may have contributed to the frequent mismatch between NQS rating and scale scores, as well as the low and minimally different mean scale scores across the NQS designations. For instance, settings rated *exceeding* on NQS scored from 1.29 (inadequate quality) to 6.71 (excellent quality) on ECERS-E, and from 1.17 to 6.70 on SSTEW. While there was a general and overall association between these quality indices, at least when considered within 2 years of each other, there were also cases of disagreement between how centres would be characterised in terms of their quality. This suggests that while NQS ratings may be well suited to ensuring sector-wide quality improvements, quality rating scales may be required in order to discriminate well between the highest levels of quality provision, consequent developmental benefits for children and to inform families of the potential benefits of their child's individual provision.

This study provides useful findings to help understand, interpret and support Australia's NQS program of assessment, across states given that the three states here showed similar patterns of results. First, while structural aspects of quality are important, especially in relation to settings showing lower levels of quality, separating the quality aspects related to children's outcomes may yield useful additional information for continued quality improvement. This, of course, requires evaluating the aspects of NQS most highly associated with children's later outcomes. In relation to the process aspects of quality already incorporated in NQS, it may be prudent to update and include more recent understandings about what is important for child development. It is also important to consider content, competence and confidence in areas like emergent mathematics, science and exploration. As an alternative to updating NQS, and in line with the actions of some US states and UK local authorities, research measurements (including quality rating scales) may be well suited to augmenting inspection and monitoring systems—especially for centres already meeting national minimum standards of quality. If governments are committed to an evidence-based approach to quality, this appears to be a useful step forward.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- ABS (Australian Bureau of Statistics). (2008). *Information Paper: An Introduction to Socio-Economic Indexes for Areas (SEIFA)* (cat. No. 2039.0). Retrieved August 16, 2016, from <http://www.abs.gov.au/ausstats/abs@.nsf/mf/2039.0>.
- Australian Children's Education & Care Quality Authority. (2017a). Australian Children's Education and Care Quality Authority. Accessed 7 June, 2017, from <http://www.acecqa.gov.au/national-quality-framework/the-national-quality-standard>.

- Australian Children's Education & Care Quality Authority. (2017b). Guide to the National Quality Standard. Accessed 6, 2017, from <http://files.acecqa.gov.au/files/National-Quality-Framework-Resources-Kit/NQF-Resource-03-Guide-to-NQS.pdf>.
- Australian Children's Education & Care Quality Authority (2018) Guide to the National Quality Framework. Accessed October, 2018, from <https://www.acecqa.gov.au/sites/default/files/2018-09/Guide-to-the-NQF.pdf>.
- Department of Children and Youth Affairs (DCYA). (2017). *Early childhood Ireland*. Dublin: Department of Children and Youth Affairs (DCYA).
- Department of Education Employment and Workplace Relations. (2009). *Belonging, being and becoming: The Early Years Learning Framework (EYLF) for Australia*. Canberra: Australian Government.
- Donabedian, A. (1980). *Explorations in quality assessment and monitoring* (Vol. 1), The definition of quality and approaches to its assessment Ann Arbor, MI: Health Administration Press.
- Education Review Office. (2010). *Quality in early childhood services*. Wellington: ERO.
- Ho, D., Campbell-Barr, V., & Leeson, C. (2010). Quality improvement in early years settings in Hong Kong and England. *International Journal of Early Years Education*, 18, 243–258.
- Howard, S. J., Siraj, I., Melhuish, E. C., Kingston, D., Neilsen-Hewett, C., de Rosnay, M., et al. (2018). Measuring interactional quality in pre-school settings: Introduction and validation of the Sustained Shared Thinking and Emotional Wellbeing (SSTEW) scale. *Early Child Development and Care*. <https://doi.org/10.1080/03004430.2018.1511549>.
- Howes, C., Burchinal, M., Pianta, R. C., Bryant, D., Early, D., Clifford, R., et al. (2008). Ready to learn? Children's pre-academic achievement in prekindergarten. *Early Childhood Research Quarterly*, 23, 27–50.
- Kingston, D. (2017) A mixed methods study: Evaluating the impact of a bespoke professional development based on an analysis of existing quality in one local authority. PhD Dissertation. London: UCL. Accessed 26 February, 2018, from <http://discovery.ucl.ac.uk/1551541/>.
- Kingston, D., & Melvin, J. (2012). Quality and effectiveness in working with children and young people. In N. Edmond & M. Price (Eds.), *Integrated working with children and young people supporting development from Birth to Nineteen*. London: Sage.
- Mathers, S., Singler, R., & Karemaker, A. (2012). *Improving quality in the early years: A comparison of perspectives and measures*. London: Nuffield Foundation.
- Melhuish, E. C., Sylva, K., Sammons, P., Siraj-Blatchford, I., Taggart, B., & Phan, M. (2008). Effects of the Home Learning Environment and preschool center experience upon literacy and numeracy development in early primary school. *Journal of Social Issues*, 64, 95–114.
- Melhuish, E., Ereky-Stevens, K., Petrogiannis, K., Ariescu, A., Penderi, E., Rentzou, K., et al. (2015). A review of research on the effects of Early Childhood Education and Care (ECEC) upon child development. CARE Project; Curriculum Quality Analysis and Impact Review of European Early Childhood Education and Care (ECEC). *Technical Report*. European Commission. Accessed 7 June, 2017, from <http://ecec-care.org/resources/publications/>.
- Moss, P., & Dahlberg, G. (2008). Beyond quality in early childhood education and care—Languages of evaluation. *New Zealand Journal of Teachers' Work*, 5, 3–12.
- New Zealand Ministry of Education. (2017). More children in early childhood education. New Zealand Ministry of Education. Accessed 5 September, 2017, from <https://education.govt.nz/ministry-of-education/government-education-initiatives/better-public-services/more-children-in-early-childhood-education/>.
- O'Connell, M., Fox, S., Hinz, B. & Cole, H. (2016) *Quality Early Education for All*, Mitchell Institute Report No. 01/2016, Victoria
- Organisation for Economic Co-operation and Development. (2006). *Starting Strong II: Early Childhood Education and Care*. Paris: Organisation for Economic Co-operation and Development.
- Organisation for Economic Co-operation and Development. (2012). *Starting Strong III: A quality toolbox for early childhood education and care*. Paris: Organisation for Economic Co-operation and Development.
- Parker, I. (2013). *Early developments: Bridging the gap between evidence and policy in early-years education*. London: Institute for Public Policy Research.
- Penn, H. (1996). Who Influences Policy in Early Childhood Services? *International Journal of Early Years Education*, 4, 5–17.
- Pianta, R. C. (2012). *Handbook of early childhood education*. London: Guildford Press.

- Plantenga, J. (2011). Local providers and loyal parents: Competition and consumer choice in the Dutch childcare market. In E. Lloyd & H. Penn (Eds.), *Childcare Markets: can they deliver an equitable service?*. Bristol: The Policy Press.
- Sabol, T. J., Soliday Hong, S. L., Pianta, R. C., & Burchinal, M. (2013). Can rating pre-K programs predict children's learning? *Science*, *341*, 845–846.
- Siraj, I., & Kingston, D. (2015). *An independent review of the Scottish Early Learning and Childcare (ELC) workforce and Out of School Care (OSC) workforce*. Edinburgh: Scottish Government.
- Siraj, I., & Mayo, A. (2014). *Social class and educational inequality. The impact of parents and schools*. Cambridge: Cambridge University Press.
- Siraj, I., Kingston, D., & Melhuish, E. (2015). *Assessing Quality in Early Childhood Education and Care. Sustained Shared Thinking and Emotional Wellbeing (SSTEW) Scale for 2–5 year olds provision*. London: IOE UCL Press.
- Siraj, I., Cheeseman, S., Kingston, D., & Lysaght, P. (2017a). *Quality Interactions Study (QIS) report: Working with DoE preschools to strengthen the quality of intentional & relational pedagogy*. Sydney: New South Wales Department of Education.
- Siraj, I., Kingston, D., Neilsen-Hewett, D., Howard, S., Melhuish, E., de Rosnay, M., et al. (2017b). *Fostering Effective Early Learning: A review of the current international evidence considering quality in early childhood education and care programmes—in delivery, pedagogy and child outcomes*. Sydney: New South Wales Department of Education. Accessed 9 June, 2017, from <https://education.nsw.gov.au/media/ecec/pdf-documents/FEEL-Study-Literature-Review-Final.pdf>.
- Siraj, I., Howard, S. J., Neilsen-Hewett, C., Melhuish, E. C., Kingston, D. de Rosnay, M., Duursma, E., & Luu, B. (2018). *Fostering Effective Early Learning (FEEL) study*, NSW Department of Education, Australia [DECEAR-15-35].
- Siraj-Blatchford, I., Sylva, K., Muttock, S., Gilden, R., & Bell, D. (2002). *Researching Effective Pedagogy in the Early Years (REPEY)—DfES Research Report 356*. London: Department for Education publishing.
- Slot, P. L., Leseman, P., Verhagen, J., & Mulder, H. (2015). Associations between structural quality aspects and process quality in Dutch early education and care settings. *Early Childhood Research Quarterly*, *33*, 64–76.
- Sylva, K., Siraj-Blatchford, I., & Taggart, B. (2003). *ECERS-E: The Early Childhood Environment Rating Scale Curricular Extension to ECERS-R*. Stoke on Trent: Trentham Books.
- Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2004). *The effective provision of pre-school education (EPPE) project: Final report*. Nottingham, UK: Department for Education.
- Sylva, K., Siraj-Blatchford, I., Taggart, B., Melhuish, E., & Sammons, P. (2006). Capturing quality in early childhood through environment rating scales. *Early Childhood Research Quarterly*, *21*, 76–92.
- Sylva, K., Melhuish, E., Sammons, P., Siraj, I., Taggart, B., Smees, R., et al. (2014). *Students' educational and developmental outcomes at age 16. Effective Pre-school, Primary and Secondary Education (EPPSE) 3-16 Project*. London: Department for Education.
- Taylor, C., Thorpe, K., Nguyen, C., Adams, R., Ishimine, K., Ferguson, A., et al. (2016). *The E4Kids study: Assessing the effectiveness of Australian early childhood education and care programs*. Melbourne: Melbourne University.
- The Early Years Foundation Stage (EYFS). (2012). *Pack*. London: Early Education.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Iram Siraj's research focuses on understanding the relationship between quality, pedagogy and early assessment on young children's developmental outcomes. She works in transdisciplinary teams on longitudinal studies to assess the long-term impact of the quality of pre-school and home environments on children from differing backgrounds.

Steven J. Howard's research investigates early child development, as well as the educational implications of these changing capacities. This includes research to identify effective approaches to supporting children's development that are low-cost and flexible, as well as identifying ways to monitor children's developmental progress that are accessible, acceptable and actionable by those who spend their time working with young children.

Denise Kingston teaches across undergraduate and graduate courses within the Department of Education and Social Work. She is a qualified teacher and Educational Psychologist and her current research interests focus on supporting and improving the quality of early childhood education and care through evidence-based professional development.

Cathrine Neilsen-Hewett is Director of the Early Years at University of Wollongong, and has particular expertise in development in early childhood, early childhood pedagogy and practice, and staff training and childhood socialization.

Edward C. Melhuish is an international expert in early child development and the impact of home learning and pre-school experiences on children's later outcomes. He also has vast experience linking child development with social policy, and is a consultant to the OECD, WHO, and EU.

Marc de Rosnay is Academic Director at Early Start, University of Wollongong, and has specific interest and expertise in social, emotional, and communicative development in the early years.