

RESEARCH

Open Access



# How stable is program quality in child care centre classrooms?

Petr Varmuza<sup>\*</sup> , Michal Perlman and Olesya Falenchuk

<sup>\*</sup>Correspondence:  
petr.varmuza@mail.utoronto.ca  
Ontario Institute for Studies  
in Education/University  
of Toronto, 252 Bloor Street  
West, Toronto, ON M5S 1V6,  
Canada

## Abstract

In the Early Childhood Education and Care (ECEC) sector there is a move to reduce oversight costs by reducing the frequency of quality assessments in providers who score highly consistently across time. However, virtually nothing is known about the stability of ECEC quality assessments over time. Using a validated measure of overall classroom quality, we examined stability of quality in a sample of over 1000 classrooms in licensed child care centres in Toronto, Canada over a 3-year period. Multilevel mixed-effects linear regression analyses revealed substantial instability across all types of ECEC centres, although publicly operated centres were somewhat more stable and tended to have higher quality scores. We also found substantial variance between classrooms within ECEC centres. None of the structural, child/family and neighbourhood characteristics we examined were significantly related to stability of quality ratings. The lack of stability found in our sample does not support the use of a risk-based approach to quality oversight in ECEC. Large within centre classroom quality variance suggest that all classrooms within a centre should be assessed individually. Furthermore, classroom level scores should be posted when scores are made public as part of accountability systems. Future research should, in addition to administrative data used in our study, explore how factors such as educator training, participation in program planning, reflective practices and ongoing learning might improve stability of quality over time.

## Introduction

A substantial and growing number of children attend Early Childhood Education and Care (ECEC) services (Friendly et al., 2018; Government of Canada, 2019; Kamerman & Gatenio-Gabel, 2007; Laughlin, 2013; Sinha, 2014). Jurisdictions in the US, Canada and elsewhere are proactively expanding ECEC access both through licensed ECEC services and introduction of full-day kindergarten with wrap-around before and after school programs (White, 2017). At the same time, public spending on these programs is increasing around the globe (OECD, 2017). However, any rapid expansion of the supply of ECEC spaces poses challenges including ensuring safe and developmentally appropriate environments, an ample supply of qualified teaching staff, and consistent, high program quality.

To try to ensure minimal standards, responsible authorities have at least some licensing mechanism in place for ECEC services. With increased demand for ECEC

services and public investment in these services, many US states/local authorities have adopted some version of a Quality Rating and Improvement System (QRIS). The goals of QRISs are to achieve better program quality and accountability and to provide information to current and potential ECEC users (Boller et al., 2015; Cannon et al., 2017; The Build Initiative & Child Trends, 2018; Tout et al., 2010). Both licensing and QRISs involve quality assessments that are conducted at the classroom and/or program level at regular intervals. However, quality assessments are labour intensive as they require non-trivial amount of time for training and administration. Reducing their frequency is one way to contain costs.

Program quality is generally defined as consisting of structural (e.g., educator/child ratios) and process dimensions (quality of educator/child interactions) (Early et al., 2007; Phillipsen et al., 1997). Structural quality indicators are generally easier to measure and regulate (Layzer & Goodson, 2006; Slot, 2018). Structural quality is thought to set the stage for the processes (e.g., educator/child interactions) that children experience directly (Bigras et al., 2010; Melhuish & Gardiner, 2019; NICHD Early Child Care Research Network, 2002; Phillipsen et al., 1997; Slot, 2018). Structural quality includes contextual factors such as funding and payment to educators which impact stability of the workforce and, therefore, influence the kinds of interactions children experience. These are thought to set the stage for process quality which includes the quality of interactions children experience with peers and adults. Thus, process quality captures children's direct experiences in care and some research suggests that process quality is the most important aspect of quality in terms of supporting children's development (Curby et al., 2009; Downer et al., 2010; Hamre et al., 2014; Mashburn et al., 2008). It is worth noting that there are other definitions that consider quality of ECEC to be a more "relative" concept. For example, Pence and Moss (1994) define quality in ECEC as "a constructed concept, subjective in nature and based on values, beliefs, and interest, rather than an objective and universal reality (p. 172)". While this is a thought-provoking definition, in this paper we focus on the more pragmatic definition of structural/process indicators since these indicators are often used by government and other stakeholder for quality monitoring purposes (OECD, 2015).

The goal of this paper is to examine the stability of classroom quality scores across time to inform whether it is possible to reduce the frequency of assessments without reducing the effectiveness of accountability systems. We also examine a number of ECEC characteristics that may predict stability in quality over time. We do this using the Assessment for Quality Improvement (AQI), which is an observational measure of overall classroom quality that is used as part of the QRIS implemented in Toronto, Canada. The AQI is a measure of global quality that captures both structural and process quality indicators. Below we describe the rationale, focusing on frequency of assessments, behind various ECEC accountability mechanisms including QRISs as implemented across the US, Toronto, Canada and examples from Europe, Asia and Australia. We then describe the potential predictors of stability included in this study. We conclude this introduction with the research questions examined in this study.

### Accountability mechanisms in the ECEC sector

While the consensus about the need for quality and for monitoring of ECEC services is clear (OECD, ), as describe as described above, definitions of quality vary. Furthermore, the actual initiatives undertaken by individual jurisdictions greatly vary. Licensing requirements usually cover structural components, such as staffing ratios, qualifications, group sizes as well as health and safety requirements. However, the possession of a license generally means having met the minimum acceptable standard of service at the time of the inspection. In many countries outside USA and Canada, additional requirements may also include program planning, curriculum implementation, financial and human resource management and working conditions (OECD, 2015). Unlike in most developed countries, the practices in USA and Canada appear to be different as a result of the de facto separation of mandatory licensing from ongoing program assessment and supports.

Compared to other developed countries, including Canada, licensing regulations that define the minimum acceptable standards in most US states tend to be weaker (Karoly, 2014; Perlman et al., 2019). According to Child Care Aware (2013) in 31 out of 50 states the minimum qualification for a lead ECEC educator was a high school diploma or less; 5 years later the number decreased marginally to 29 (Whitebook et al., 2018). However, it is important to note many individual ECEC services or types of programs (e.g., publicly funded preschool programs) require much higher levels of educational qualification. To improve quality standards, many state and local administrations have added QRISs as another layer of voluntary oversight. QRISs involve in-depth assessments of ECEC providers that serve multiple goals, including giving ECEC providers useful quality improvement feedback, using ratings for accountability purposes and enabling parents to make more informed decisions for their children. While participation may be required to maintain eligibility for state funding, in most states less than 50% of child care centres participate in their local QRIS (The Build Initiative & Child Trends, 2018). Minimum ratings for a QRIS tier/level are often established by local consensus. In many instances, an observational assessment that captures, among other things, the quality of educator/child interaction, is required only for the higher QRIS tiers (The Build Initiative & Child Trends, 2018). These are often captured using one of the measures in the suite of measures referred to as Environmental Rating Scales (ERSs, e.g., the Early Childhood Environmental Rating Scale-3 and the Infant Toddler Environmental Rating Scale-Revised) (Sylva et al., 2006; Vermeer et al., 2016).

ERS scores, together with structural indicators, such as educator levels of formal education, are used to create a composite score reflecting different tiers of quality that are usually reported on a scale of one to five stars. The QRIS ratings in the USA are usually valid for 1–3 years, with 3 years being the most common duration (The Build Initiative & Child Trends, 2018). According to the OECD Starting Strong IV study (2015) the frequency of monitoring practices can vary from several times per year (Luxemburg and Mexico) to annual (e.g., Japan, Chile, Mexico, Netherlands), to every 2 or 3 years (e.g., Ireland, France, Korea, Belgium).

Many researchers (Bassok et al., 2016; Blau, 2007; Gorry & Thomas, 2017; Hotz & Xiao, 2011; Loeb et al., 2004; Scarr, 1998) argue that regulations unnecessarily increase the burden on operators and reduce access to ECEC services for low-income families in

particular. Blau (2007) also finds that regulations negatively affect ECEC workers' wages, possibly contributing to lower program quality. This view argues for reducing regulatory burden as a way of reducing costs to users (Gorry & Thomas, 2017) and promoting creativity in service delivery. One way that has been proposed for reducing the burden of oversight is to decrease the frequency of oversight visits.

The cost of assessments is deemed to be too high and different state administrations have explored different ways to reduce those costs. For example, some allow for selecting a smaller number of classrooms to be assessed within a centre (Tout et al., 2010). However, this approach is problematic because the limited research that exists on this issue suggests that there is substantial variability in quality between classrooms within centres (Karoly et al., 2013; Perlman et al., 2019; Sabol et al., 2019). Another approach that is gaining in popularity is reducing the frequency of assessments. An increasing number of US jurisdictions engage in "differential monitoring", both in licensing and in QRISs, based on previously assessed levels of compliance (The National Center on Early Childhood Quality Assurance, 2015). This means longer lags between visits for providers who have a consistently high track record in terms of regulatory assessments. Similarly, in many countries including Australia, England, and New Zealand, the frequency of assessment visits is based on previous ratings and risk assessments (OECD, 2015).

The differential monitoring approach assumes that ECEC providers maintain a stable level of quality across time, and therefore, a longer interval between visits of high functioning ECEC settings will still accurately capture ECEC provider performance. While this may allow a rechanneling of oversight resources to higher risk programs, to our knowledge this assumption of stability, which underlies increasing lags between assessments for licensing and QRIS's has yet to be empirically tested.

### **Toronto's QRIS**

In Ontario, where the current study took place, the provincial government provides the majority of funding as well as the regulatory framework including control over staffing, group sizes and issuance of licenses for all types of ECEC services. The government has recently introduced a tiered licensing system based on the approaches advocated by the National Administration for Regulatory Administration (NARA).

Upper tier municipalities (counties and regions) are designated as local service managers of the ECEC system in Ontario, of which licensed centre-based care is only one component. The City of Toronto is the largest municipality in Canada with approximately 170,000 children below the age of six. At the beginning of 2014, the City of Toronto reported having 41,646 licensed child care spaces in 852 centres. Approximately 70% of these centres were eligible to provide care for subsidized children and represent the frame of our study. Child care subsidies (24,264 in January 2014) (City of Toronto Children's Services, 2014) are portable vouchers accepted by any eligible provider. To be eligible for a subsidy, parents must be either working or studying full-time and meet income requirements. The value of the voucher is set by the actual costs of delivering care reported by each specific ECEC provider. This means that in Toronto, the cost of child care should not play any role in selection of the child care centre and organizational characteristics of the service provider for parents receiving a child care subsidy.

The City of Toronto operates a QRIS that, together with other components, includes mandatory annual AQI assessments for all programs eligible for placement of subsidized children. The AQI was developed in Toronto. To be eligible to provide care for children receiving a subsidy, providers must score a minimum of 3 out of 5 on the AQI. The AQI was developed to be relatively efficient (it takes approximately 90 min to administer, which is far more efficient than 3–5 h of other frequently used ERSs). The preschool version of the AQI that is used in the current study has been shown to significantly correlate to other measures of ECEC classroom quality (e.g., it is correlated to the ECERS-R at 0.61,  $p < 0.01$ ). Following each assessment, the results are reviewed with the centre supervisor and posted on the centre's notice board. Centre staff are encouraged to discuss the visit results as part of ongoing program development. Results are also used as a basis for providing tailored quality improvement supports provided by City of Toronto coaches/consultants. Finally, results are posted online to enable parents to use ratings when selecting ECEC providers for their children. Given the policy context of this research (i.e., the trend towards reducing the frequency of quality assessments) we set out to examine whether certain centre and classroom characteristics may predict higher/lower stability over time.

#### **Factors that might impact stability of classroom quality**

In the absence of published literature on the stability of quality ratings, and because the assumption of the tiered licensing and quality assessments links quality to stability, we selected a number of age group and centre characteristics for inclusion in the analysis because we hypothesize that they may mediate quality, as well as the stability of quality rating. These include centre type (auspice and organizational status), neighbourhood status, percentage of service delivered by educators with a qualification in Early Childhood Education, hourly wages for early childhood educators and centre supervisors, program size, presence of other age groups, proportion of subsidized children and proportion of subsidized children who come from single-parent families. These variables have been examined in the context of studies on quality; here we focusing on their implications for the stability of quality over time. The rationale for inclusion of each of the covariates is discussed in more detail below.

In Canada, as in many other countries, there is an ongoing debate about the differences in program quality related to program auspice<sup>1</sup> (Brennan et al., 2012; Cleveland, 2008; Cleveland & Krashinsky, 2009; Penn, 2011; LLOYD & Penn, 2012; Mitchell, 2012; Morris & Helburn, 2000; Moss, 2012; Paull, 2012; Sosinsky & Kim, 2013; Sosinsky et al., 2007). Although the majority of researchers find that there are some differences in quality, often they are confounded by the market conditions (Bassok et al., 2016; Cleveland & Krashinsky, 2009) or neighbourhood conditions (Small et al., 2008; Sosinsky et al., 2007). Based on previously published research in Canada (Cleveland, 2008, Cleveland et al., 2008) we expect non-profit centres to show greater stability in terms of quality over time. Furthermore, provincial and municipal funding and system management policies in Toronto are applied differentially to commercial, non-profit and public programs. As

---

<sup>1</sup> We use the term "auspice" to define the ownership of the centre and the term "centre type" to identify whether the centre belongs to a multi-site operation or is operated as a single site entity.

a result, we used the type of centre is used as a control (stratifying) variable. We also distinguish between single centre and multiple centre organizations, expecting that multi-site organizations may be more stable over time since they may have better policies and procedures in place to streamline their operations and service delivery. We hypothesize that larger, multi-site operations would exhibit smaller standard deviations in AQI scores across time and, presumably, higher AQI scores.

Neighbourhood status has been found to relate to the type and quality of available child care (Bassok & Galdo, 2016; Burchinal et al., 2008; Hatfield et al., 2015; Vandenberg & Lazzari, 2014). The actual mechanisms of neighbourhood effects are not clear and are empirically difficult to prove (Galster, 2012; Galster et al., 2011; Ham et al., 2012; Manley et al., 2013). Some have even questioned whether they exist within the Canadian context at all (Oreopoulos, 2008). Nonetheless, given past findings, we control for neighbourhood effects by deploying the Child and Family Inequity Score (CFIS), described in detail below.

Educator qualification is an important contributor to the quality of care (Arnett, 1989; Manning et al., 2019; Whitebook et al., 2001). Ontario's child care regulations require at least one staff with a minimum 2-year degree or diploma in early childhood education (ECE) in every preschool classroom. However, many ECEC programs operate with a higher number of ECEs in all classrooms. We expect that the higher levels of training will be associated with greater stability. Similarly, better remunerated staff provide higher levels of care (Schleicher, 2019). We expect that higher rates of pay would be associated with higher rates of stability in quality across time.

From the authors' ongoing work with City of Toronto administrative data on an unrelated study, we learned that majority of subsidized children who start care as infants remain enrolled at least until they reach the kindergarten age. Child care centre programs that serve infants in effect generate their toddler and preschooler enrollment from the children who started in the centre as infants. Given what the literature suggests about the positive effects of early child care enrolment (Sylva et al., 2011), we hypothesize that, besides the individual child effects, the preschool programs potentially composed of children who were enrolled as infants should exhibit higher and more stable quality scores.

Every centre in our database accepts subsidized children as a condition of its contract with the City of Toronto. However, the proportions of subsidized children and the proportions of subsidized children who live in single-parent families vary greatly between centres, primarily as a result of the neighbourhood status. Centres with a higher proportion of children from low-income, predominantly single-parent families experience a higher level of enrolment turnover due to changes in families' subsidy eligibility. We theorize that these family characteristics will negatively affect the quality and stability of quality in centres with a high proportion of subsidized and single-parent families.

### **The research questions**

Our primary goal was to examine the stability of preschool classroom quality over time. Our secondary goal is to test whether specific classroom characteristics might predict

stability, enabling the identification of classrooms that might be good targets for less frequent and, therefore, less costly oversight through longer lags between quality assessments.

To do this we examine classroom level stability in quality over a 3-year period. We use the population of centres that are part of the City of Toronto's QRIS using the AQI. Specifically, our research questions are:

1. Are classroom quality scores stable across the 3-year period?
2. Are quality scores in some programs more stable than others over time?
  - a. Is stability related to program quality? Specifically, we expect that higher quality classrooms would be more stable over time.
  - b. If higher quality classrooms are more stable, what are their distinguishing characteristics? We expect that covariates associated with high quality (specifically auspice, neighbourhood, proportion of educators with ECE degrees, hourly wage rate, presence of infant classrooms, proportion of children receiving a subsidy, and proportion of subsidized children in single-parent families) would also be associated with stability.
  - b. Are higher quality classrooms sufficiently stable to reduce the frequency of assessments?

## Methods

### Data

The present study includes all preschool child care classrooms in the City of Toronto that were part of the City of Toronto's QRIS every year between 2014 and 2016. The municipal government maintains an extensive administrative database which includes budget information, staffing, public and subsidy fees, as well as data on adherence to performance standards measured using the AQI. Any program interested in delivering subsidized child care in Toronto must be part of the City's QRIS. As a result, our study consists of the entire population of centres that took part in this system. The author requested and obtained permission to access the data in raw form from the City of Toronto Children's Services. In 2014, the 1st year of the data utilized in this study, 70.3% of preschool programs, 73.9% of toddler programs and 82.8% of infant programs in Toronto participated in the City's QRIS. The remaining centres either did not wish to provide access to subsidized children or were deemed ineligible by either City of Toronto policy on restricting growth of the commercial child care sector or a Council approved Child Care Service Plan.

The final study frame consisted of 501 centres with 1019 preschool classrooms. Table 1 shows the number of centres and classrooms for which 3 years of data are available for the analysis presented in this article. The table distinguishes between different types of operators based on auspice and whether the operator owns multiple sites.

### Measures

#### *AQI—the assessment for quality improvement initiative*

The preschool version of the AQI is a measure of overall quality consisting of 31 items (see the list of individual items in Appendix 1). Similar to the ECERS-3, the scoring

**Table 1** Study frame—child care centres and number of rooms by operator type

	Commercial		Non-profit		Municipal	Total
	Single	Multi	Single	Multi		
Centres	50	43	162	205	41	501
	10.0%	8.6%	32.3%	40.9%	8.2%	100.0%
Classrooms	88	94	358	414	65	1019
	8.6%	9.2%	35.1%	40.6%	6.4%	100.0%

**Table 2** AQI Scores by centre type and year

Centre type	Rooms	Year 0		Year 1		Year 2	
		M	SD	M	SD	M	SD
Commercial single	88	4.01	0.37	3.97	0.33	4.07	0.37
Commercial multi	94	4.11	0.41	4.12	0.33	4.28	0.38
Non-profit single	358	4.00	0.43	3.97	0.37	4.10	0.38
Non-profit multi	414	4.07	0.41	4.06	0.39	4.12	0.38
Municipal	65	4.39	0.24	4.39	0.23	4.47	0.26
All centre types	1019	4.06	0.41	4.05	0.38	4.15	0.39

system requires that all sub-item scores on one level meet the standard before moving to the next higher level. The original validation study (Perlman & Falenchuk, 2010) found a one factor solution in which the mean of all individual items is taken as the reported score. It also found that a factor comprised of the mean score of items related to the quality of teacher–children interactions can be calculated and used as a stand-alone factor. The Spearman correlation between the measure and ECERS-R was  $r=0.61$   $p<0.01$ . The Spearman correlations for the CLASS subscales of Emotional Support, Classroom Organization and Instructional Support were  $r=0.39$   $p<0.01$ ,  $r=0.36$   $p<0.01$  and  $r=0.47$   $p<0.01$ , respectively. The current version of the AQI is measured on a five-point scale where scores of 1 and 2 represent inadequate quality, a score of 3 meets the City’s minimum standards to maintain a service contract, and individual item scores of 4 and 5 exceed minimum expectation. Any items with scores below three identified during the assessment are subject to a remediation order and further sanctions if the identified problems persist.

AQI assessments are conducted unannounced annually by trained observers who are randomly assigned to ECEC centres. All classrooms within a centre are assessed by the same rater. Raters’ interrater reliability is established every 4 months and assessors must achieve 80% or higher agreement with gold standard expert ratings. The average inter-rater agreement for 2014, 2015 and 2016 was 94%, 96% and 92%, respectively. During this period individual raters’ percent agreement scored ranges from 81 to 100%. For publication on the City of Toronto website the AQI scores are aggregated to the age group level; however, classroom level scores were available for analysis in our data set. In this paper we focus on the cross-sectional and longitudinal characteristics of total AQI ratings.

Descriptive statistics for the cross-sectional data for each of 3 years is presented in Table 2. The overall mean AQI value for all classrooms has increased over 3 years by



0.09, or approximately 5% of the acceptable range of 3–5. The standard deviations have changed only marginally over time; given that the effective range of acceptable scores is between three and five, we interpret the standard deviations as large. Finally, we note that the mean scores differ between the individual centre types as does the rate of change over type.

### **Independent and control variables**

#### ***Centre type***

Because provincial and municipal funding and system management policies are applied differentially to commercial, non-profit and public programs, the type of centre is used as a control (stratifying) variable. It is defined by the combination of auspice and the number of centres operated by individual service providers. Auspice is defined as commercial, non-profit, and publicly operated. Any organization that comprises three or more preschool sites is categorized as multi-site operator.

#### ***The Child and Family Inequity Score (CFIS)***

CFIS is an index developed by the City of Toronto in co-operation with representatives of community and post-secondary institutions. The index is composed of the following items: incidence of children in low-income families, female education unemployment rate, lack of affordable housing and proportion of families with English as a second language. The individual items are assigned weights by consensus and calculated for each of Toronto's 140 neighbourhoods. In this sample CFIS scores range from – 1.5 to 2.14 with the lower values representing more child and family friendly and affluent neighbourhoods.

#### ***Centre and staff characteristics***

The number of preschool spaces in each centre, as well as the presence of infant and kindergarten age groups are used as indicators of size of each centre. Staff characteristics include the percentage of care hours delivered by Early Childhood Educators (ECE) with a minimum 2-year post-secondary degree, trained staff (ECE), and average ECE and centre supervisor hourly wage. This information is only available at the program level (i.e., the same score is used for all classrooms that serve preschool aged children within a centre).

#### ***Child and family demographics***

The proportion of subsidized children in this study is extracted from the administrative database for 2015 to match the centre profile and neighbourhood data. We include the proportion of subsidized children in one parent families as an additional proxy of low family income. As with the case for educator level variables, this information was only available at the program level.

### **Analytic approach**

We begin with a cross-sectional analysis of mean AQI scores for each of 3 years by individual centre type. Because of the unequal group sizes and variances, where appropriate

**Table 3** Study covariates by centre type

Covariates	All centres (N = 501)	Commercial		Non-profit		Municipal (N = 41)
		Single (N = 50)	Multi (N = 43)	Single (N = 162)	Multi (N = 205)	
ECE hours (%)	71.25	59.53	62.22	74.43	71.99	77.86
ECE hourly wage	22.52	17.90	18.16	23.46	22.21	31.85
SUP hourly wage	34.16	28.26	30.89	36.14	31.85	50.71
# preschool spaces	31.05	32.70	40.21	30.27	29.16	31.83
Infant (%/100)	0.36	0.50	0.59	0.25	0.33	0.71
Kindergarten (%/100)	0.62	0.38	0.57	0.77	0.63	0.14
Subsidized (%)	44.97	64.87	67.23	31.70	42.69	73.40
1 parent family (%)	53.22	63.16	62.80	47.88	51.97	63.02
Neighbourhood CFIS	0.18	0.55	0.56	- 0.08	0.16	0.46

we analyze individual centre types separately as opposed to including centre type as a predictor.

To answer our first research question, we use Stata 15 (*Stata Statistical Software: Release 15, 2017*) software to build a growth model specifying random slope and random intercept to partition within and between variances over 3-year period. The null model consists of annual observations of AQI scores for each preschool classroom within a given centre. We test the model fit by examining the interclass correlations and residuals. We then built a separate null model for each centre type to assess their individual within and between classroom variances.

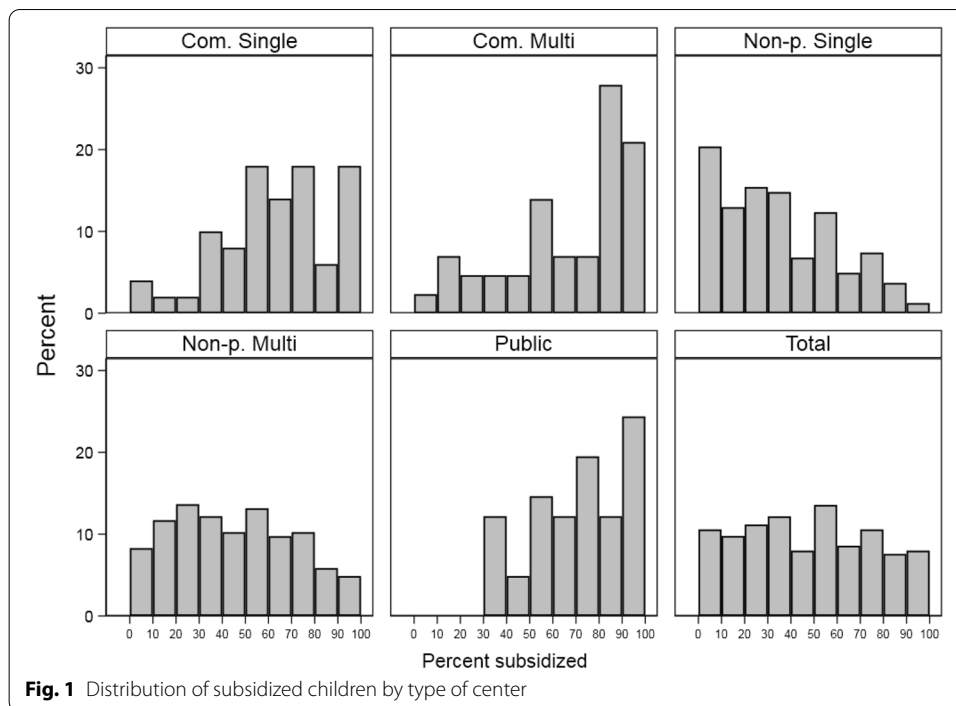
To analyze further the characteristics of classrooms with stable scores we compute the maximum difference between the highest and lowest classroom score and define as stable those classrooms with differences between zero and  $-1$  standard deviation. We then conduct a visual analysis to investigate the distribution of stable classrooms across the range of AQI scores.

An OLS regression using the maximum score difference as a dependent variable with a full set of independent variables described above is performed to answer RQ 2-a. To validate our conclusion, we also execute separately for each centre type a logistic regression with AQI stability as the dependent variable. Finally, to answer question RQ 2-c we calculate the percentages of classrooms in each centre type that retained a position in the top 75th percentile of AQI scores in each year.

## Results

### Description of the data

Table 3 provides the average values for the covariates taken from year zero (2014) of the study, with the exception of the CFIS which was based on 2016 Canadian census data. As can be seen in Table 3, the average values of individual covariates vary substantially for the different types of centres (see Appendix 2 for statistical comparisons of their



means). These differences along with findings that the different centre types had unequal variances indicate that centre types need to be examined separately in this sample.

Pairwise comparison of means finds that CFIS for municipal and commercial centres of either type are not significantly different from each other. Non-profit centres, both single and multi-site, are primarily located in neighbourhoods with significantly lower CFIS values (i.e., more affluent neighbourhoods). At the same time, the mean CFIS of multi-site, non-profit centres is significantly higher than that of single site non-profits ( $F(1,365) = 8.49, p < 0.01$ ). Although there is a large correlation ( $r = 0.582, p < 0.001$ ) between CFIS and proportion of subsidized children with the preschool age group, the proportion of children receiving a subsidy is included because it reflects the characteristics of the actual children in each centre. The histograms presented in Fig. 1 demonstrate the substantial differences in distributions of subsidized children, which are not apparent when all centre types are combined. Appendix 2 presents the results of mean comparisons for all covariates, including significance levels adjusted for unequal group sizes and unequal variances.

Three-hundred and seventy-one (36.4%) out of the 1019 classrooms are in centres that provide service to the infant age group. The presence of younger age groups is associated with different levels of AQI scores in classrooms that serve preschool aged children. Specifically, using a *t* test with unequal variances option we find that these classrooms ( $M = 4.19, SD = 0.41$ ) were rated significantly higher than classrooms in centres without infants ( $M = 3.99, SD = 0.40, t(1) = -7.22, p < 0.001$ ).

**Can classroom AQI scores be aggregated in centres with multiple preschool classrooms?**

The way the AQI is used as part of the City of Toronto’s accountability system involves quality aggregation across classrooms that serve children of the same age within centres.

**Table 4** Comparison of AQI means by type and year with adjustment for unequal variances and sample sizes with Tamhane’s T2

	Year 0				Year 1				Year 2			
	Diff	SE	t	adj. p>t	Diff	SE	t	adj. p>t	Diff	SE	t	adj. p>t
CM vs. CS	0.10	0.06	1.76	n.s	0.14	0.05	2.92	<0.05	0.21	0.06	3.86	<0.001
NS vs. CS	-0.01	0.06	-0.22	n.s	0.00	0.04	0.01	n.s	0.04	0.04	0.83	n.s
NM vs. CS	0.05	0.04	1.22	n.s	0.09	0.04	2.27	n.s	0.06	0.04	1.26	n.s
M vs. CS	0.38	0.05	7.74	<0.001	0.41	0.05	9.15	<0.001	0.40	0.05	7.87	<0.001
NS vs. CM	-0.11	0.05	-2.32	n.s	-0.14	0.04	-3.58	<0.01	-0.18	0.04	-4.05	<0.001
NM vs. CM	-0.05	0.05	-1.02	n.s	-0.05	0.04	-1.34	n.s	-0.16	0.04	-3.69	<0.001
M vs. CM	0.28	0.05	5.34	<0.001	0.27	0.04	5.99	<0.001	0.19	0.05	3.71	<0.001
NM vs. NS	0.06	0.03	2.11	n.s	0.09	0.03	3.26	<0.05	0.02	0.03	0.66	n.s
M vs. NS	0.39	0.04	10.49	<0.001	0.41	0.03	11.81	<0.001	0.37	0.04	9.60	<0.001
M vs. NM	0.32	0.04	9.15	<0.001	0.32	0.03	9.35	<0.001	0.35	0.04	9.31	<0.001

CS commercial single, CM commercial multi, NS non-profit single, NM non-profit multi, M municipal

These aggregations make assumptions about homogeneity in classroom quality that have received only limited attention from researchers (Károly et al., 2013; Pauker et al., 2018). To determine whether it is appropriate to combine across preschool classrooms within a centre, we examine the variability in classroom quality within centres. In the 345 centres that had more than one preschool classroom, the mean range between the lowest and highest score is 0.32 with a standard deviation of 0.25. A decomposition of variance into between classrooms and across centres shows only a moderate level of intraclass correlation between classrooms (ICC = 0.597, SE = 0.030 CI 95(0.536, 0.654)). Not surprisingly, the mean range of values increases with the number of rooms. However, even centres with only two rooms have an average range of 0.26. A range of 0.32 represents 16% of the acceptable range between 3.00 and 5.00.

**Descriptive results: AQI scores by type of centre and year**

Given this level of heterogeneity in the quality scores of classrooms within centres, we analyzed stability across time for individual classrooms. Average AQI scores were comparable across centre type with one exception. Using a method that adjusts for unequal variances and unequal sample sizes, a pairwise comparison of AQI means between individual centre types (Table 4) reveals that only publicly operated centres score consistently higher than the other centre types. Multi-site commercial operators show increasing, statistically significant differentiation from the non-profit and single commercial centres over the study period.

**RQ 1—Are classroom quality scores stable across the 3-year period?**

To avoid potential problems of different group sizes and unequal variances, we have estimated the multilevel model separately for each individual centre type as well as for the entire sample. The results of all estimation including the intraclass correlations are presented in Table 5.

The fixed effects part of the model across all classrooms shows an intercept of 4.05 with a slope of 0.04; in the random effects part of the model the variances for year and intercepts are displayed together with the residual variance. The total variance of

**Table 5** AQI growth model for all classrooms, and stratified by centre type; with intraclass correlations on age group and classroom levels

	Commercial single AQI	Commercial single AQI	Commercial multi AQI	Non-profit single AQI	Non-profit multi AQI	Municipal AQI
Fixed effects						
Slope	0.041*** (0.007)	0.026 (0.023)	0.083*** (0.019)	0.050*** (0.012)	0.027** (0.010)	0.039* (0.019)
Intercept	4.052*** (0.015)	3.982*** (0.041)	4.083*** (0.048)	3.975*** (0.027)	4.063*** (0.023)	4.375*** (0.029)
Random effects						
var(Age group—Intercept)	0.073*** (0.006)	0.045*** (0.012)	0.072 (.)	0.068*** (0.010)	0.071*** (0.009)	0.014*** (0.006)
var(Classroom—Slope)	0.008*** (0.003)	0.008*** (0.008)	0.000 (.)	0.013*** (0.005)	0.007*** (0.004)	0.005*** (0.005)
Intercept	0.010*** (0.005)	0.004** (0.008)	0.000 (.)	0.026*** (0.010)	0.005*** (0.008)	0.000 (0.009)
Cov(Slope Intercept)	− 0.008* (0.003)	− 0.005 (0.008)	0.000 (.)	− 0.019 (0.006)	− 0.005 (0.005)	− 0.001 (0.006)
var(Residual)	0.077*** (0.003)	0.074*** (0.009)	0.067 (.)	0.084*** (0.005)	0.077*** (0.005)	0.037*** (0.007)
N	3057	264	282	1074	1242	195
<b>Level</b>	<b>ICC</b>	<b>ICC</b>	<b>ICC</b>	<b>ICC</b>	<b>ICC</b>	<b>ICC</b>
Age group	0.457	0.363	0.518	0.382	0.465	0.275
Age group Classroom	0.518	0.393	0.518	0.528	0.498	0.279

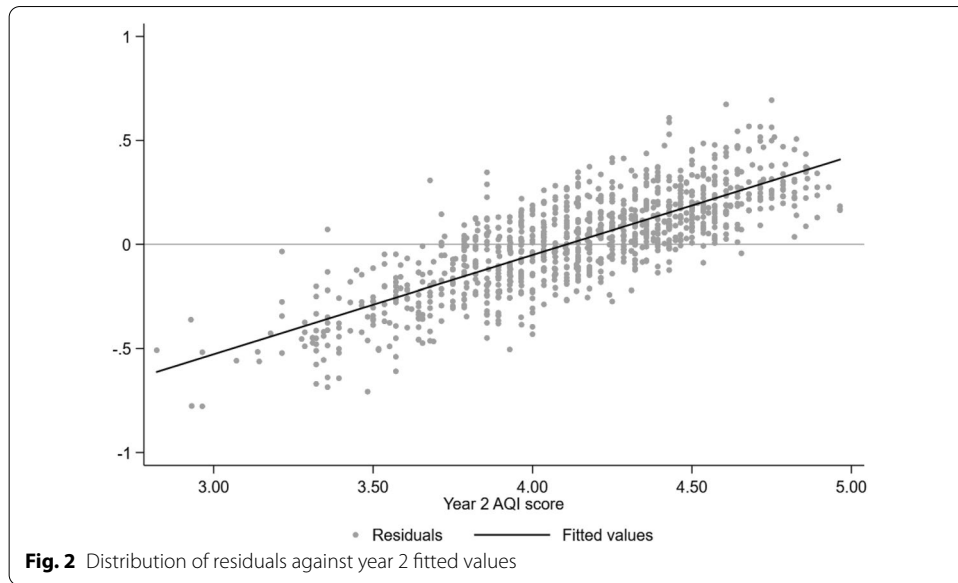
Standard errors in parentheses  
\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

random effects is used as a numerator in the calculation of intraclass correlation (ICC) with the denominator being the total variance of random effects plus residual variance. The ICC values of 0.457 for the age group level and 0.518 for the combined age group and classroom levels suggest that slightly more than half of the total variance originates between classrooms, while the remainder represents the within classroom variance. In other words, the chance of accurately predicting the next score is only slightly better than 50%,<sup>2</sup> thus allowing us to answer the Research Question #1 in negative. Focusing on individual centre types reveals a range of annual growth from 0.026 AQI in single-site commercial centres to 0.083 in multi-site centres. Similarly, the joint age group-classroom intraclass correlations range from as low as 0.279 in public programs to 0.528 in single-site non-profit programs further confirming that large within classroom variances negate the possibility of safely predicting the AQI scores in succeeding years.

The magnitude of the residual variances suggests possible issues with the estimation itself. To begin with, although the residuals are approximately normally distributed, a

<sup>2</sup> However, a note of caution is required in interpreting the ICC, especially in cases where the random effects variances (Rabe-Hesketh & Skrondal, 2008) are very low as is the case regarding publicly operated centres where both random effects variance are low (especially for the AQI score) and residual variance is less than one half of that for all classrooms combined.

As can be seen from Table 4, the AQI scores of publicly operated centres have a substantially smaller standard deviation than other programs. This, combined with higher intercept scores, leads to lower variances of the intercept; in this case 0.0160 for public centres vs 0.0837 for all centres combined. Thus, despite having the lowest variance of residuals, it also has the lowest ICC scores.



**Table 6** Average differences between lowest and highest AQI scores of individual classrooms by centre type

Centre type	M	SD	Min	Median	Max	IQR
Commercial single	0.47	0.29	0.03	0.39	1.61	0.39
Commercial multi	0.46	0.25	0.04	0.39	1.16	0.39
Non-profit single	0.53	0.30	0.03	0.50	1.93	0.41
Non-profit multi	0.49	0.26	0.04	0.46	1.43	0.39
Municipal	0.35	0.18	0.01	0.36	0.75	0.32
All centre types	0.49	0.27	0.01	0.46	1.93	0.40

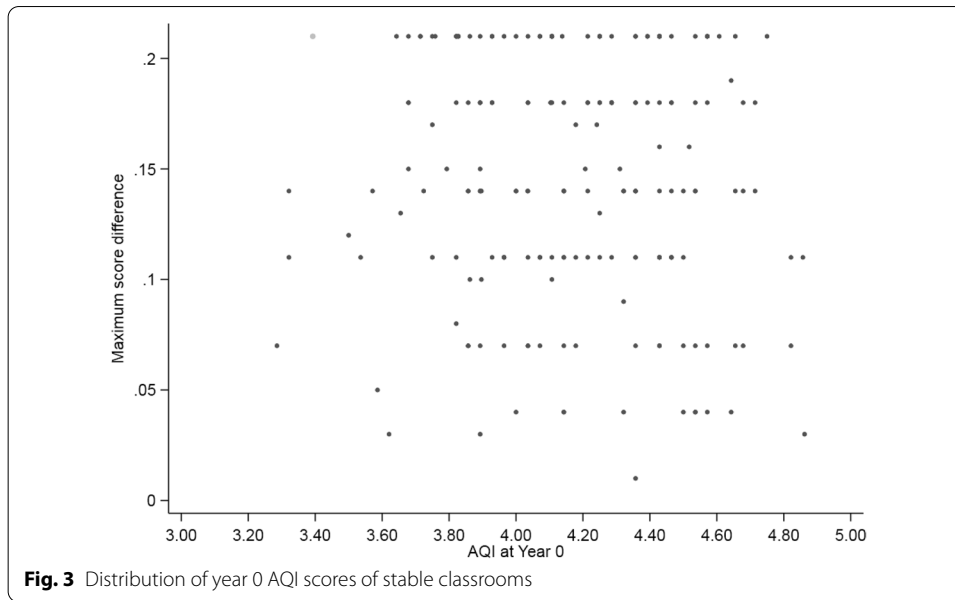
IQR refers to interquartile range

plot of residuals against fitted values (Fig. 2) reveals a lack of random distribution around zero. Positive residuals indicate that the fitted value underestimates the actual value, while a negative residual indicates overestimation. Because there is an upper bound on positive residual values that is equal to fitted value plus residual being less or equal to 5.00, the plot shows a reduction in positive residuals around the 4.50 level of the AQI. At the same time, the strong correlation between residuals and fitted values ( $r=0.778$ ,  $p<0.001$ ) suggests that the linear estimation process does not represent well the actual AQI trajectories.

**RQ 2—Are quality scores in some programs more stable than others over time?**

**RQ 2a—Is stability related to program quality?**

To determine the stability of scores we calculate the absolute difference between the highest and lowest scores for each classroom in the 3-year period. The mean difference between the highest and lowest score for all rooms is 0.49 with standard



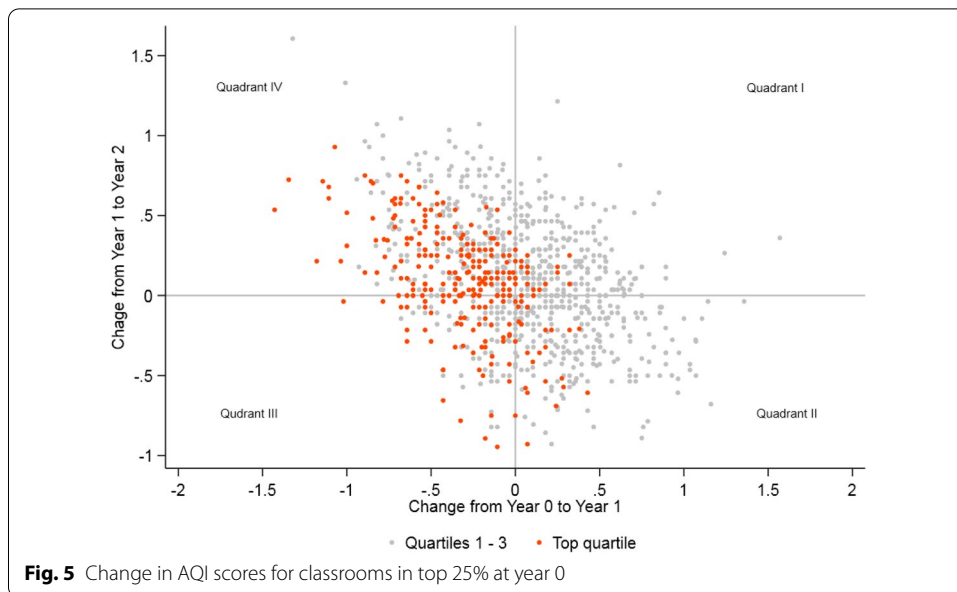
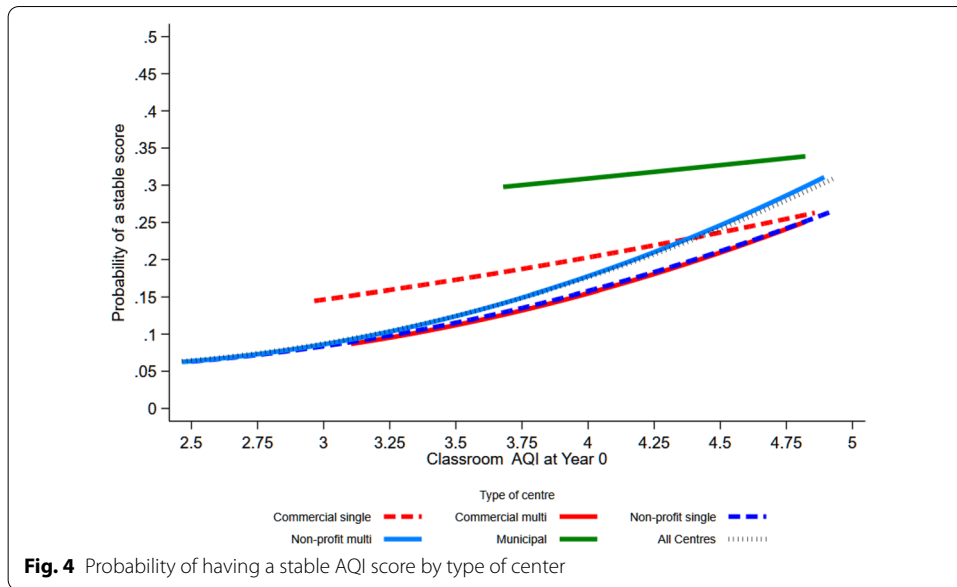
deviation of 0.27 (Table 6) with municipal centres’ mean difference being significantly lower than that of the other centre types. All classrooms with difference values lower than minus 1 SD (0.22) are then deemed to be “stable”. Using this approach, 192 or 18.8% of the 1019 classrooms are deemed stable. This percentage varies by type of centre, ranging from a low of 16.2% for non-profit single site centres to a high of 32.2% for the municipal programs.

Notably, the absolute difference values tell us nothing about the direction of change; of all 1019 classrooms only 13% improved their AQI score in each year, while 7% declined every year. The remaining 80 experienced a variety of patterns that included growth, reduction or stability of AQI scores.

As shown in Fig. 3, there is no easily discernible relationship between the maximum score difference and the AQI at year 0 for stable classrooms; in other words, classrooms with stable scores can be found across the whole range of AQI scores. However, the Pearson correlation between the year 0 score and the maximum difference is weak at  $r = -0.2703$ ,  $p < 0.0001$ . We confirm this finding by plotting the results of separate logistic regressions with stability as the dependent variable and AQI score at year 0 for each centre type as the independent variable. While the probability of any classroom having stable scores increases with their initial (year 0) scores, it never reaches even a 40% level (Fig. 4).

**RQ 2b—If higher quality classrooms are more stable, what are their distinguishing characteristics?**

Multiple regression analyses for the full sample as well as for individual centre types are used to test whether the difference in classroom scores over the 3-year



period could be predicted from the covariates used in this study. The results reveal that the program characteristics explain only 2% of the variance ( $\text{adj. } R^2 = 0.0233$ ,  $F(9,1006) = 3.69$ ,  $p < 0.001$ ) for the full sample and non-significant results for individual centre types. The full results are presented in the Appendix 4.

**RQ 2c—Are high quality classrooms sufficiently stable to reduce the frequency of assessments?**

To answer this question, we identify the 25% of classrooms that scored highest on the AQI in year 0 and track the changes in their AQI score to year 1 and then from year



**Table 7** Stability of AQI scores of highest ranked classrooms in year 0

Classrooms	CS	CM	NS	NM	M	Total
All classrooms	88	94	358	414	65	1019
Top 25% in Year 0	14	29	73	107	31	254
% of all classrooms	15.9%	30.9%	20.4%	25.8%	47.7%	24.9%
Top 25% every year	2	12	12	27	17	70
% of all classrooms	2.3%	12.8%	3.4%	6.5%	26.2%	6.9%
% of Top 25%	14.3%	41.4%	16.4%	25.2%	54.8%	27.6%

CS commercial single, CM commercial multi, NS non-profit single, NM non-profit multi, M municipal

1 to year 2. The classrooms in the top AQI quartile in year 0 are identified as orange dots in the scatterplot presented in Fig. 5. The AQI scores of a large proportion of the top classrooms dropped from year 0 to year 1—Quadrants II and IV. From year 1 to year 2 many of the classrooms made up some, if not all, of the previous year's drop in AQI scores—Quadrant IV. Classrooms whose AQI scores dropped in both, year 1 and year 2 relative to their score in year 0 are in Quadrant III.

However, because we are interested in the stability of the scores for the purpose of reducing the frequency of oversight assessments, it is illustrative to focus on the number of classrooms that manage to retain their membership in the top 25% in each of 3 years. The distribution of top scoring classrooms in year 0 among the types of centres is shown in Table 7. The proportions of high scoring programs range from 15.9% for single site commercial operators to 47.7% for municipally operated programs. Over the 3-year span less than 7% of all 1,019 classrooms manage to remain in the top quartile of AQI scores.

The percentage of programs that consistently maintain their top ranking is shown in the last row of the same table. Of all the 254 classrooms that are in the top 25% in year 0, only 27.6% (70) remain in that group in each of the following 2 years. Across the types of centres, the rate ranges from low of 14.3% in the single site commercial programs to a high of 54.8% for the municipal programs. Even at the much higher retention level in municipal programs is not sufficient to exempt these programs from annual assessment.

## Discussion

### Stability of classrooms scores over a 3-year period

One of the main motivations for this study was to empirically test the degree to which program quality is stable over time. Using multi-level modelling we establish that the within classroom variances are almost as high as the between classroom variances. In practical terms this means that it is almost impossible to accurately predict the next year's score from the current year.<sup>3</sup> When focusing on centres that scored in the 75th percentile or higher in year 0, we find a substantial differences in their ability to maintain the high ranking according to the centre type (Table 7); however, even within the highest scoring municipal sector, only 54.8% managed to retain that ranking over the 3-year period.

<sup>3</sup> In post-hoc analyses we explored different models such as latent class trajectories without any improvement to our ability to predict quality scores in the succeeding years.

We also find stable scores at all levels of quality (Fig. 3). This, of course, is problematic as classrooms at lower levels of quality should focus on continuing improvement. Finding stability at all levels of quality also helps to explain the lack of associations between stability and the structural characteristics usually associated with program quality.

On the other hand, the programs at the high end of the scale are expected to maintain their ratings over time. We have defined high quality scores as belonging to the 75th percentile or higher in year 0. Contrary to our expectations, less than 28% of classrooms managed to remain in the top category every single year in the 3-year period. This finding leads us to reject the suggestion that the frequency of assessments can be reduced on the basis of belonging to the top scoring programs. In a post hoc analysis we employ the same approach to analyze classrooms with scores in the 90th percentile and above; only 15% of classrooms manage to maintain their place in that category in each of 3 years. Therefore, even among exceptionally strong programs, instability is very high.

There are significant differences based on the type of the centre in the membership in the top-quality groups as well as in the rate of remaining in the group over 3 years; we address these differences below.

### **Centre type**

Based on the differences in centre characteristics, we expected differences in levels of quality and stability, albeit tempered by the strong system management role provided by City of Toronto. Compared to both types of commercial centres, on average, the non-profit sector pays significantly higher wages, operates with a higher proportion of ECE trained staff, is located in more affluent neighbourhoods and serves a lower proportion of subsidized children (Appendix 2). Staff characteristics of the public sector are comparable to the non-profit sector, while the child and family demographics, and neighbourhood characteristics are similar to those of the commercial sector centres. A comparison of the quality scores across the five centre types showed no significant differences between the commercial and non-profit centres in year 0 and significant difference between the public centres and the rest. The public centre advantage remains in the following 2 years while the commercial multi-site sector improved its score enough to separate itself from the other centre types.

To understand the similarity of scores between commercial and non-profit providers it is important to note that the commercial centres in this sample have all been part of the City's QRIS for decades. The relatively tight enforcement of the standards and supports within the City's QRIS may explain the relatively high performance of the commercial centres (Cleveland, 2008). Since the early 1990's the City of Toronto has had a policy of not entering into any new contracts with commercial operators as well as eliminating profit as a component of approved operating budgets. All ECEC operators with a purchase of service contract with the City of Toronto are under a non-distribution constraint, making them effectively "commercial, entrepreneurial" non-profits (Bushouse, 1999; Hansmann, 1980). At the same time, the operators officially designated as non-profit receive higher operating grants making it possible for them to pay higher wages or hire more than the legislated minimum of educators with relevant educational backgrounds. The variability in AQI scores within the commercial and non-profit sectors

suggests that more attention should be paid to supporting quality of service rather than whether the centre falls into commercial or non-profit category.

We categorize centres into single and multiple site operators to explore whether being a part of a larger organization contributes to higher consistency of practices and standards of operations. A more consistent operation would be expected to exhibit smaller standard deviations and, presumably, higher AQI scores. However, a closer examination of several multi-site agencies reveals no consistent relationship between higher level of stability and quality scores. Although the public centres exhibit substantially higher level of stability (Table 4 and Fig. 4), they are still well below the rate that would allow for reduced frequency of quality assessments.

#### **Classroom or age group level?**

Aggregation of scores across classrooms to the centre level is generally adopted in many QRIS systems; including the one currently operated by City of Toronto's QRIS. However, we find that in many cases there are substantial differences ( $M=0.32$ ,  $SD=0.25$ ) between individual classrooms in centres. Because all of a centre's classrooms are assessed by one, and only one, trained observer, the issue of inter-rater reliability does not apply in any given year. Nevertheless, above and beyond the analysis presented in this paper, the aggregation of individual scores has some serious implications. First, it potentially misleads users about the program quality of their child's classroom; in this study the difference of 0.32 on the AQI scale represents a 16% difference. Second, any substantial difference should give rise to questions about program supervision and management practices. Finally, it supports the recommendation that, rather than a sample within the centre, all classrooms should be assessed on a regular basis.

#### **Cost of child care quality assessments in City of Toronto**

The cost of assessments is covered by the municipality and at the time of data collection was approximately 8 cents per child per day (A. Hepditch, personal communication, January 8, 2020). This represents less than 0.02% of a median price of a preschool space in Toronto.

#### **Limitations**

This study suffers from several limitations. In limitation is that the study data come from the City of Toronto which is a high demand market area (i.e., a seller's market); this is demonstrated by the growth in commercial and non-profit operators who opt to remain outside the subsidy system, primarily in affluent areas of the city. Results from this study are primarily generalizable to localities with a similar market profile and level of oversight and program support.

The administrative data used in this study were collected for accountability and performance improvement of programs that serve various proportions of subsidized and full-fee children. Although these programs represent over 70% of preschool services in Toronto, no conclusions should be drawn about the quality of the non-funded programs or programs that were established in 2015 or later. However, it is important to note the 70% coverage rate is relatively high compared to other studies.

Another limitation of the administrative data we use is that information that would be valuable in exploring program quality and stability is simply not available. Under the provisions of protection of privacy legislation, the municipality is not allowed to collect data for the purpose of determining eligibility for subsidized child care related to ethnic background, language spoken at home and parent education. Instead we rely on neighbourhood information (CFIS) which is more distal to the actual service than the parents of children enrolled in the centre. Another limitation of our data is that the three data-points available to us did not enable us to test for curvilinear patterns in that data. It is important that future studies including longer term follow-up of quality test for such patterns.

The information about the proportion of ECEs in each classroom, their hourly pay, and the number of children receiving a subsidy and who came from single-parent households is only available as an aggregate across all preschool aged classrooms in each centre.

Finally, the issue of human error in measurement is an important one to consider. Although the program observers are regularly tested, and had high levels of interrater agreement exceeding 90% across raters and time, an interrater agreement rate of 100% is generally not feasible. This means that some level of disagreement exists between raters and across time and this likely explains some of the fluctuations in scores observed in this study. One way we reduced the role of measurement error/noise in our analyses is that we did not consider very low levels of fluctuations as reflective of instability. In general, the question of the levels of fluctuation that are significant requires further study and it will be important to include measures of child wellbeing as a way to determine which levels of fluctuations in quality are meaningful.

Two different findings are noteworthy and merit further investigation. First, there is no evidence that program auspice can be used in predicting stability of quality scores. Even though as a group, publicly operated programs had significantly higher AQI scores and lower variability, a small number of those programs fell below the expected quality and stability range.

Second, further investigation is needed to identify factors underlying the stability of quality, or the lack thereof. The covariates—centre, user and neighbourhood characteristics—used in this study shed little light on this question. Future research should examine these variables at the classroom level as well as explore additional variables such as educator training, staff retention, participation in program planning, reflective practices and ongoing learning might improve stability of quality over time. Because the data available to us consist of annual assessments we were not able to test the extent to which quality varies within the year.

## **Conclusion**

We set out to provide empirical evidence of the stability of quality ratings over a 3-year period, and to investigate whether quality assessments can be carried out on a less frequent than annual basis. Our findings do not support such a change. In fact, our findings suggest that the frequency of assessments should not be reduced because attaining a high score in any given year is not a guarantee of doing so again in subsequent years. Furthermore, the chance of remaining in the top quartile scoring group over the 3-year period is less than 28%. In addition, if we accept that one of the purposes of any QRIS is

to provide evidence of effective program intervention and supports, then such evidence has to be available in a timely manner. Reducing the frequency of independent assessments will only make it difficult to identify critical issues, devise corrective strategies and provide required program supports. Finally, if the information resulting from QRIS or, more specifically quality assessments, is to be useful to parents in making their child care choices, then it has to be current. Together, these highlight the need to maintain the frequency of quality assessments, conducted annually at a minimum, as part of ECEC quality oversight regimes.

## Appendices

### Appendix 1. AQI preschool items

Item	Name
1.	Daily and visual schedules
2.	Program plan
3.	Learning experiences
4.	Indoor physical environment
5.	Displays
6.	Sensory, science and nature
7.	Art
8.	Books
9.	Language and literacy
10.	Music and accessories
11.	Physical play learning experiences
12.	Blocks and construction
13.	Cognitive and manipulative
14.	Dramatic play
15.	Electronic media usage <sup>a</sup>
16.	Toileting and diapering procedures
17.	Meal and/or snack times
18.	Equipment required for eating and seating
19.	Cots and bedding <sup>a</sup>
20.	Health and safety
21.	Toys and equipment washing
22.	Staff and children's hygiene
23.	Transitions
24.	Attendance verification
25.	Positive atmosphere
26.	Supervision of children
27.	Foster children's experience
28.	Supporting the Development of Self-Esteem
29.	Behaviour guidance

Item	Name
30.	Supporting the development of communication skills
31.	Extending children's learning

<sup>a</sup>Not included in analysis due to large amount of missing data

**Appendix 2. Comparison of covariate means by type of centre, controlling for unequal group sizes and variances with Tamhane's T2**

Comparing	Percent ECE hours				ECE hourly wage			
	Diff	Std. Err	Tamhane's T2		Diff	Std. Err	Tamhane's T2	
			t	adj. p > t			t	adj. p > t
CM vs. CS	2.460	2.779	0.89	0.991	0.179	0.518	0.35	1.000
NS vs. CS	14.856	2.496	5.95	0.000	5.384	0.532	10.12	0.000
NM vs. CS	12.268	2.388	5.14	0.000	4.331	0.517	8.37	0.000
M vs. CS	18.280	2.101	8.70	0.000	13.906	0.458	30.33	0.000
NS vs. CM	12.396	2.323	5.34	0.000	5.205	0.366	14.24	0.000
NM vs. CM	9.807	2.207	4.44	0.000	4.152	0.343	12.09	0.000
M vs. CM	15.820	1.892	8.36	0.000	13.727	0.246	55.75	0.000
NM vs. NS	- 2.588	1.837	- 1.41	0.825	- 1.053	0.365	- 2.88	0.041
M vs. NS	3.424	1.444	2.37	0.173	8.522	0.276	30.91	0.000
M vs. NM	6.012	1.248	4.82	0.000	9.575	0.245	39.01	0.000

Comparing	Supervisor hourly wage				Size of preschool program			
	Diff	Std. Err	Tamhane's T2		Diff	Std. Err	Tamhane's T2	
			t	adj. p > t			t	adj. p > t
CM vs. CS	2.003	1.539	1.30	0.888	6.136	2.881	2.13	0.309
NS vs. CS	7.352	1.324	5.55	0.000	- 3.943	2.005	- 1.97	0.417
NM vs. CS	3.318	1.302	2.55	0.125	- 3.995	1.907	- 2.10	0.333
M vs. CS	22.428	1.257	17.84	0.000	- 1.590	2.328	- 0.68	0.999
NS vs. CM	5.349	1.067	5.01	0.000	- 10.079	2.508	- 4.02	0.002
NM vs. CM	1.315	1.040	1.26	0.907	- 10.131	2.430	- 4.17	0.001
M vs. CM	20.425	0.983	20.78	0.000	- 7.726	2.773	- 2.79	0.066
NM vs. NS	- 4.034	0.683	- 5.91	0.000	- 0.052	1.274	- 0.04	1.000
M vs. NS	15.076	0.592	25.45	0.000	2.353	1.846	1.27	0.901
M vs. NM	19.111	0.543	35.21	0.000	2.405	1.739	1.38	0.848

Comparing	Proportion of centres with infants				Proportion of centres with kindergarten			
	Diff	Std. Err	Tamhane's T2		Diff	Std. Err	Tamhane's T2	
			t	adj. p > t			t	adj. p > t
CM vs. CS	0.078	0.105	0.75	0.998	0.145	0.102	1.43	0.820
NS vs. CS	- 0.184	0.080	- 2.30	0.218	0.303	0.077	3.95	0.002
NM vs. CS	- 0.119	0.079	- 1.51	0.768	0.217	0.075	2.88	0.050
M vs. CS	0.227	0.101	2.24	0.243	- 0.198	0.084	- 2.35	0.193
NS vs. CM	- 0.262	0.085	- 3.09	0.029	0.158	0.086	1.84	0.516
NM vs. CM	- 0.197	0.084	- 2.36	0.198	0.071	0.085	0.85	0.994
M vs. CM	0.149	0.105	1.42	0.824	- 0.343	0.093	- 3.70	0.004
NM vs. NS	0.065	0.049	1.31	0.878	- 0.087	0.052	- 1.68	0.628

Comparing	Proportion of centres with infants				Proportion of centres with kindergarten			
	Diff	Std. Err	Tamhane's T2		Diff	Std. Err	Tamhane's T2	
			t	adj. p > t			t	adj. p > t
M vs. NS	0.411	0.080	5.11	0.000	- 0.502	0.064	- 7.80	0.000
M vs. NM	0.346	0.079	4.36	0.001	- 0.415	0.062	- 6.64	0.000
Comparing	Percent subsidized				Percent subsidized with single parents			
	Diff	Std. Err	Tamhane's T2		Diff	Std. Err	Tamhane's T2	
			t	adj. p > t			t	adj. p > t
CM vs. CS	3.928	5.271	0.75	0.998	0.785	3.605	0.22	1.000
NS vs. CS	- 27.971	3.855	- 7.26	0.000	- 12.887	3.146	- 4.10	0.001
NM vs. CS	- 17.448	3.788	- 4.61	0.000	- 9.443	2.912	- 3.24	0.016
M vs. CS	8.770	4.640	1.89	0.473	0.685	3.665	0.19	1.000
NS vs. CM	- 31.898	4.529	- 7.04	0.000	- 13.672	3.252	- 4.20	0.001
NM vs. CM	- 21.376	4.472	- 4.78	0.000	- 10.228	3.026	- 3.38	0.012
M vs. CM	4.843	5.213	0.93	0.988	- 0.099	3.757	- 0.03	1.000
NM vs. NS	10.523	2.662	3.95	0.001	3.444	2.461	1.40	0.831
M vs. NS	36.741	3.777	9.73	0.000	13.572	3.319	4.09	0.001
M vs. NM	26.219	3.708	7.07	0.000	10.128	3.097	3.27	0.017
Comparing	Child and Family Inequity Index							
	Diff	Std. err	Tamhane's T2					
			t	adj. p > t				
CM vs. CS	0.007	0.166	0.04	1.000				
NS vs. CS	- 0.634	0.123	- 5.15	0.000				
NM vs. CS	- 0.400	0.121	- 3.31	0.014				
M vs. CS	- 0.091	0.162	- 0.56	1.000				
NS vs. CM	- 0.641	0.139	- 4.61	0.000				
NM vs. CM	- 0.406	0.137	- 2.97	0.042				
M vs. CM	- 0.098	0.174	- 0.56	1.000				
NM vs. NS	0.235	0.080	2.92	0.036				
M vs. NS	0.543	0.135	4.03	0.002				
M vs. NM	0.309	0.133	2.33	0.212				

CS commercial single site, CM commercial multiple site, NS non-profit single site, NM non-profit multiple site, M municipal

**Appendix 3. Pairwise correlations of study covariates**

	(1) ECE hours (%)	(2) ECE hourly wage	(3) Super- visor hourly wage	(4) Preschool spaces	(5) Infants present	(6) KG present	(7) Subsidy (%)	(8) With single parent (%)	(9) Inequity score
(1) ECE hours (%)	1.000								
<i>p</i>	0.000								
(2) ECE hourly wage	0.321	1.000							
<i>p</i>	0.000	0.000							
(3) Super- visor hourly wage	0.243	0.735	1.000						
<i>p</i>	0.000	0.000	0.000						
(4) Pre- school spaces	-0.092	-0.093	0.105	1.000					
<i>p</i>	0.039	0.038	0.019	0.000					
(5) Infants pre- sent	-0.069	0.061	0.194	0.135	1.000				
<i>p</i>	0.124	0.176	0.000	0.002	0.000				
(6) KG pre- sent	0.016	-0.078	-0.080	-0.095	-0.417	1.000			
<i>p</i>	0.725	0.081	0.074	0.034	0.000	0.000			
(7) Sub- sidy (%)	-0.213	-0.030	0.022	-0.113	0.221	-0.230	1.000		
<i>p</i>	0.000	0.500	0.618	0.011	0.000	0.000	0.000		
(8) With single parent (%)	-0.099	-0.047	-0.056	-0.113	0.006	-0.033	0.369	1.000	
<i>p</i>	0.027	0.296	0.209	0.012	0.893	0.466	0.000	0.000	
(9) Ineq- uity score (CFIS)	-0.112	-0.112	-0.074	-0.031	0.139	-0.073	0.582	0.270	1.000
<i>p</i>	0.012	0.012	0.100	0.496	0.002	0.101	0.000	0.000	0.000

**Appendix 4. Linear regressions of study covariates on maximum AQI difference over 3 years**

	(1) All centres	(2) Com. single	(3) Com. multi	(4) Non-p. single	(5) Non-p multi	(6) Public
ECE hours (%)	0.0002 (0.0006)	-0.0042* (0.0021)	0.0004 (0.0026)	0.0015 (0.0010)	-0.0003 (0.0008)	0.0133 (0.0152)
ECE hourly wage	-0.0014 (0.0029)	0.0153 (0.0157)	-0.0101 (0.0221)	-0.0021 (0.0056)	-0.0011 (0.0048)	0.1370 (0.1140)



	(1)	(2)	(3)	(4)	(5)	(6)
	All centres	Com. single	Com. multi	Non-p. single	Non-p multi	Public
SUP hourly wage	− 0.0023 (0.0016)	− 0.0069 (0.0053)	− 0.0100 (0.0052)	0.0000 (0.0031)	− 0.0010 (0.0025)	− 0.0154 (0.0138)
# preschool spaces	− 0.0013 (0.0007)	− 0.0014 (0.0028)	− 0.0007 (0.0020)	− 0.0005 (0.0017)	− 0.0014 (0.0012)	− 0.0042 (0.0026)
Infant (%/100)	0.0343 (0.0204)	0.0257 (0.0722)	0.0333 (0.0655)	0.0905* (0.0390)	0.0149 (0.0349)	0.1090 (0.0629)
Kindergarten (%/100)	0.0591** (0.0200)	− 0.0615 (0.0713)	0.0369 (0.0554)	0.0942* (0.0410)	0.0456 (0.0338)	− 0.0623 (0.0759)
Subsidized (%)	− 0.0006 (0.0004)	0.0001 (0.0020)	0.0015 (0.0017)	0.0003 (0.0010)	− 0.0008 (0.0007)	− 0.0033* (0.0014)
1 parent family (%)	0.0003 (0.0004)	− 0.00449* (0.0019)	− 0.0027 (0.0021)	0.0005 (0.0007)	0.0012 (0.0007)	0.0008 (0.0018)
Neighbourhood CFIS	− 0.0190 (0.0137)	0.0014 (0.0601)	0.0430 (0.0391)	− 0.0347 (0.0276)	− 0.0068 (0.0214)	− 0.0284 (0.0339)
Constant	0.594*** (0.0714)	0.968*** (0.279)	0.955 (0.518)	0.342* (0.154)	0.544*** (0.124)	− 4.010 (4.297)
Observations	1,016	88	94	355	414	65
R-sq. adjusted	0.023	0.072	0.010	0.016	0.007	0.072
F-statistic	3.693	1.748	1.099	1.657	1.308	1.555

Standard errors in parentheses \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

#### Acknowledgements

We gratefully acknowledge the comments of Jane Bertrand, Donna Lero and Linda White on earlier drafts of this article.

#### Authors' contributions

The authors are responsible for the reported research. All authors have participated in conceptualization and design, analysis and interpretation of data and in drafting or revising of the manuscript. All authors read and approved the final manuscript.

#### Funding

This research was carried out without external funding.

#### Availability of data and materials

Data are available from the corresponding author upon reasonable request, subject to permission from the City of Toronto Children's Services.

#### Declarations

##### Competing interests

The authors declare that they have no competing interests.

Received: 28 March 2020 Accepted: 11 October 2021

Published online: 28 October 2021

#### References

- Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10(4), 541–552. [https://doi.org/10.1016/0193-3973\(89\)90026-9](https://doi.org/10.1016/0193-3973(89)90026-9)
- Bassok, D., Fitzpatrick, M., Greenberg, E., & Loeb, S. (2016). Within- and between-sector quality differences in early childhood education and care. *Child Development*, 87(5), 1627–1645. <https://doi.org/10.1111/cdev.12551>
- Bassok, D., & Galdo, E. (2016). Inequality in preschool quality? Community-level disparities in access to high-quality learning environments. *Early Education and Development*, 27(1), 128–144. <https://doi.org/10.1080/10409289.2015.1057463>
- Bigras, N., Bouchard, C., Cantin, G., Brunson, L., Coutu, S., Lemay, L., Tremblay, M., Japel, C., & Charron, A. (2010). A comparative study of structural and process quality in center-based and family-based child care services. *Child & Youth Care Forum*, 39(3), 129–150. <https://doi.org/10.1007/s10566-009-9088-4>
- Blau, D. M. (2007). Unintended consequences of child care regulations. *Labour Economics*, 14(3), 513–538. <https://doi.org/10.1016/j.labeco.2006.01.003>

- Boller, K., Paulsell, D., Grosso, P. D., Blair, R., Lundquist, E., Kassow, D. Z., Kim, R., & Raikes, A. (2015). Impacts of a child care quality rating and improvement system on child care quality. *Early Childhood Research Quarterly*, 30, 306–315. <https://doi.org/10.1016/j.ecresq.2014.10.001>
- Brennan, D., Cass, B., Himmelweit, S., & Szebehely, M. (2012). The marketisation of care: Rationales and consequences in Nordic and liberal care regimes. *Journal of European Social Policy*, 22(4), 377–391.
- Burchinal, M., Nelson, L., Carlson, M., & Brooks-Gunn, J. (2008). Neighborhood characteristics, and child care type and quality. *Early Education & Development*, 19(5), 702. <https://doi.org/10.1080/10409280802375273>
- Bushouse, B. K. (1999). *The mixed economy of child care: An institutional analysis of nonprofit, for-profit and public enterprises* (Ph.D Thesis). Indiana University.
- Cannon, J. S., Zellman, G. L., Karoly, L. A., & Schwartz, H. L. (2017). *Quality rating and improvement systems for early care and education programs: Making the second generation better*. RAND.
- Child Care Aware. (2013). *We can do better: Child Care Aware of America's ranking of state child care center regulations and oversight*. Retrieved from [https://usa.childcareaware.org/wp-content/uploads/2015/10/wecandobetter\\_2013\\_final\\_april\\_11\\_0.pdf](https://usa.childcareaware.org/wp-content/uploads/2015/10/wecandobetter_2013_final_april_11_0.pdf).
- City of Toronto Children's Services. (2014). *Early learning and care in Toronto—January 2014*. City of Toronto.
- Cleveland, G. (2008). *If it don't make dollars, does that mean that it don't make sense? Commercial, nonprofit and municipal child care in the City of Toronto*. City of Toronto. Retrieved from <http://www.toronto.ca/legdocs/mmis/2008/cd/bgrd/backgroundfile-12677.pdf>.
- Cleveland, G., & Krashinsky, M. (2009). The nonprofit advantage: Producing quality in thick and thin child care markets. *Journal of Policy Analysis and Management*, 28(3), 440–462.
- Cleveland, G., Forer, B., Hyatt, D., Japel, C., & Krashinsky, M. (2008). New evidence about child care in Canada: Use patterns, affordability and quality. *IRRP Choices*, 14(12). <http://archive.irpp.org/choices/archive/vol14no12.pdf>.
- Curby, T. W., LoCasale-Crouch, J., Konold, T. R., Pianta, R. C., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2009). The relations of observed Pre-K classroom quality profiles to children's achievement and social competence. *Early Education & Development*, 20(2), 346. <https://doi.org/10.1080/10409280802581284>
- Downer, J., Sabol, T. J., & Hamre, B. (2010). Teacher-child interactions in the classroom: Toward a theory of within- and cross-domain links to children's developmental outcomes. *Early Education & Development*, 21(5), 699–723. <https://doi.org/10.1080/10409289.2010.497453>
- Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., Cai, K., Clifford, R. M., Ebanks, C., Griffin, J. A., Henry, G. T., Howes, C., Iriondo-Perez, J., Jeon, H.-J., Mashburn, A. J., Peisner-Feinberg, E., Pianta, R. C., Vandergrift, N., & Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: results from seven studies of preschool programs. *Child Development*, 78(2), 558–580.
- Friendly, M., Larsen, E., & Feltham, L. (2018). *Early childhood education and care in Canada 2016*. Retrieved from <https://apps.uqo.ca/LoginSigparb/LoginPourRessources.aspx?url=http://www.deslibris.ca/ID/10096600>.
- Galster, G. C. (2012). *The mechanism(s) of neighbourhood effects: Theory, evidence, and policy implications* (pp. 23–56). Springer.
- Galster, G. C., Quercia, R. G., & Cortes, A. (2011). Identifying neighborhood thresholds: An empirical exploration. *Housing Policy Debate*, 11(3), 701–732. <https://doi.org/10.1080/10511482.2000.9521383>
- Gorry, D., & Thomas, D. W. (2017). Regulation and the cost of childcare. *Applied Economics*, 49(41), 4138–4147. <https://doi.org/10.1080/00036846.2016.1276275>
- Government of Canada, S. C. (2019, April 10). *The daily—Survey on early learning and child care arrangements, 2019*. Retrieved from <https://www150.statcan.gc.ca/n1/daily-quotidien/190410/dq190410a-eng.htm>.
- Ham, M. V., Manley, D., Bailey, N., Simpson, L., & Maclennan, D. (2012). *Neighbourhood effects research: New perspectives*. Springer.
- Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher-child interactions: Associations with preschool children's development. *Child Development*, 85(3), 1257–1274. <https://doi.org/10.1111/cdev.12184>
- Hansmann, H. B. (1980). The role of nonprofit enterprise. *The Yale Law Journal*, 89(5), 835. <https://doi.org/10.2307/796089>
- 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.
- Hotz, V. J., & Xiao, M. (2011). The impact of regulations on the supply and quality of care in child care markets. *The American Economic Review*, 101(5), 1775–1805.
- Kamerman, S. B., & Gatenio-Gabel, S. (2007). Early childhood education and care in the United States: An overview of the current policy picture. *International Journal of Child Care and Education Policy*, 1(1), 23–34. <https://doi.org/10.1007/2288-6729-1-1-23>
- Karoly, L. A. (2014). *Validation studies for early learning and care quality rating and improvement systems*. RAND Education and RAND Labor and Population.
- Karoly, L. A., Zellman, G. L., & Perlman, M. (2013). Understanding variation in classroom quality within early childhood centers: Evidence from Colorado's quality rating and improvement system. *Early Childhood Research Quarterly*, 28(4), 645–657. <https://doi.org/10.1016/j.ecresq.2013.05.001>
- Laughlin, L. (2013). *Who's minding the kids?: Child care arrangements, Spring 2011*. US Department of Commerce, Bureau of the Census.
- Layzer, J., & Goodson, B. (2006). The "Quality" of early care and education settings. *Evaluation Review*, 30(5), 556–576.
- Lloyd, E., & Penn, H. (Eds.). (2012). *Childcare markets: Can they deliver an equitable service?* (1st ed.). The Policy Press, University of Bristol.
- Loeb, S., Fuller, B., Kagan, S. L., & Carrol, B. (2004). Child care in poor communities: Early learning effects of type, quality, and stability. *Child Development*, 75(1), 47–65.
- Manley, D., van Ham, M., Bailey, N., Simpson, L., & Maclennan, D. (2013). *Neighbourhood effects or neighbourhood based problems?* Springer.

- Manning, M., Wong, G. T. W., Fleming, C. M., & Garvis, S. (2019). Is teacher qualification associated with the quality of the early childhood education and care environment? A meta-analytic review. *Review of Educational Research*, 89(3), 370–415. <https://doi.org/10.3102/0034654319837540>
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., Burchinal, M., Early, D. M., & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79(3), 732–749.
- Melhuish, E., & Gardiner, J. (2019). Structural factors and policy change as related to the quality of early childhood education and care for 3–4 year olds in the UK. *Frontiers in Education*. <https://doi.org/10.3389/educ.2019.00035>
- Mitchell, L. (2012). *Markets and childcare provision in New Zealand: Towards a fairer alternative*. In E. Lloyd & H. Penn, (Eds.). (1st ed., p. 97). The Policy Press, University of Bristol.
- Morris, J., & Helburn, S. (2000). Child care center quality differences: The role of profit status, client preferences, and trust. *Nonprofit and Voluntary Sector Quarterly*, 29(3), 377–399.
- Moss, P. (2012). *Need markets be the only show in town?*. In E. Lloyd & H. Penn (Eds.). (1st ed., p. 191). The Policy Press, University of Bristol.
- NICHD Early Child Care Research Network. (2002). Child-care structure? Process? Outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science*, 13(3), 199–206. <https://doi.org/10.1111/1467-9280.00438>
- OECD. (2012). *A quality toolbox for early childhood education and care*. OECD.
- OECD. (2015). *Starting strong IV: Monitoring quality in early childhood education and care*. OECD Publishing. <https://doi.org/10.1787/9789264233515-en>
- OECD. (2017). *Public expenditure on childcare and early education services*. OECD Publishing. Retrieved from [https://www.oecd-ilibrary.org/economics/economic-policy-reforms-2017/public-expenditure-on-childcare-and-early-education-services\\_growth-2017-graph176-en](https://www.oecd-ilibrary.org/economics/economic-policy-reforms-2017/public-expenditure-on-childcare-and-early-education-services_growth-2017-graph176-en)
- Oreopoulos, P. (2008). Neighbourhood effects in Canada: A critique. *Canadian Public Policy*, 34(2), 237–258.
- Pauker, S., Perlman, M., Prime, H., & Jenkins, J. (2018). Caregiver cognitive sensitivity: Measure development and validation in Early Childhood Education and Care (ECEC) settings. *Early Childhood Research Quarterly*, 45, 45–57. <https://doi.org/10.1016/j.ecresq.2018.05.001>
- Paull, G. (2012). *Childcare markets and government intervention*. In E. Lloyd & H. Penn (Eds.). (1st ed., p. 227). The Policy Press, University of Bristol.
- Pence, A., & Moss, P. (1994). *Towards an inclusionary approach in defining quality*. Sage.
- Penn, H. (2011). Gambling on the market: The role of for-profit provision in early childhood education and care. *Journal of Early Childhood Research*, 9(2), 150–161.
- Perlman, M., & Falenchuk, O. (2010). *DOES THE CITY OF TORONTO'S MEASURE OF CHILD CARE CENTRE QUALITY WORK AS INTENDED?* Vol. Report prepared for City of Toronto Children's Services.
- Perlman, M., Howe, N., Gulyas, C., & Falenchuk, O. (2019). Associations between directors' characteristics, supervision practices and quality of early childhood education and care classrooms. *Early Education and Development*. <https://doi.org/10.1080/10409289.2019.1679006>
- Phillipsen, L. C., Burchinal, M. R., Howes, C., & Cryer, D. (1997). The prediction of process quality from structural features of child care. *Early Childhood Research Quarterly*, 12(3), 281–303.
- Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata* (2nd ed.). STATA Press.
- Sabol, T. J., Ross, E. C., & Frost, A. (2019). Are all head start classrooms created equal? Variation in classroom quality within head start centers and implications for accountability systems. *American Educational Research Journal*. <https://doi.org/10.3102/0002831219858920>
- Scarr, S. (1998). American child care today. *American Psychologist*, 53(2), 95–108.
- Schleicher, A. (2019). Policies for early learning: Work organisation and staff qualifications. In A. Schleicher (Ed.), *Educating our Youngest* (pp. 21–38). OECD. <https://doi.org/10.1787/d3abb9cb-en>
- Sinha, M. (2014). *Child care in Canada 2014*. Statistics Canada, Social and Aboriginal Statistics Division. Retrieved from <http://www.deslibris.ca/ID/244808>.
- Slot, P. (2018). *Structural characteristics and process quality in early childhood education and care* (No. 176; OECD Education Working Papers). OECD. Retrieved from <https://doi.org/10.1787/edaf3793-en>.
- Small, M. L., Jacobs, E. M., & Massengill, R. P. (2008). Why organizational ties matter for neighborhood effects: Resource access through childcare centers. *Social Forces*, 87(1), 387–414.
- Sosinsky, L. S., & Kim, S.-K. (2013). A profile approach to child care quality, quantity, and type of setting: Parent selection of infant child care arrangements. *Applied Developmental Science*, 17(1), 39–56.
- Sosinsky, L. S., Lord, H., & Zigler, E. (2007). For-profit/nonprofit differences in center-based child care quality: Results from the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development. *Journal of Applied Developmental Psychology*, 28(5–6), 390–410.
- Stata Statistical Software: Release 15*. (2017). StataCorp.
- Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., & Totsika, V. (2006). Capturing quality in early childhood through environmental rating scales. *Early Childhood Research Quarterly*, 21(1), 76–92.
- Sylva, K., Stein, A., Leach, P., Barnes, J., Malmberg, L., & The FCCC-Team. (2011). Effects of early child-care on cognition, language, and task-related behaviours at 18 months: An English study: Effects of early child-care. *British Journal of Developmental Psychology*, 29(1), 18–45. <https://doi.org/10.1348/026151010X533229>
- The Build Initiative & Child Trends. (2018). *A catalog and comparison of quality rating and improvement systems*. Quality Compendium Resources. Retrieved from <https://qualitycompendium.org/resources/qriscompendium.org-resources>.
- The National Center on Early Childhood Quality Assurance. (2015). *Trends in Child Care Center Licensing Regulations and Policies for 2014* (Research Brief #1). Retrieved from [https://childcareta.acf.hhs.gov/sites/default/files/public/center\\_licensing\\_trends\\_brief\\_2014.pdf](https://childcareta.acf.hhs.gov/sites/default/files/public/center_licensing_trends_brief_2014.pdf).

- Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G., & Boller, K. (2010). *Compendium of quality rating systems and evaluations* [Mathematica Policy Research Reports]. Mathematica Policy Research. Retrieved from <https://EconPapers.repec.org/RePEc:mpr:mprres:e3e1472c256f4bd49f20ac5b3c962ad7>.
- Vandenbroeck, M., & Lazzari, A. (2014). Accessibility of early childhood education and care: A state of affairs. *European Early Childhood Education Research Journal*, 22(3), 327–335. <https://doi.org/10.1080/1350293X.2014.912895>
- Vermeer, H. J., van Ijzendoorn, M. H., Cárcamo, R. A., & Harrison, L. J. (2016). Quality of child care using the environment rating scales: A meta-analysis of international studies. *International Journal of Early Childhood*, 48(1), 33–60. <https://doi.org/10.1007/s13158-015-0154-9>
- White, L. A. (2017). *Constructing policy change: Early childhood education and care in liberal welfare states*. University of Toronto Press.
- Whitebook, M., McLean, C., Austin, L. J. E., & Edwards, B. (2018). *Early Childhood Workforce Index—2018*. Center for the Study of Child Care Employment, University of California, Berkeley. Retrieved from <http://cscce.berkeley.edu/topic/early-childhood-workforce-index/2018/>.
- Whitebook, M., Sakai, L., Gerber, E., & Howes, C. (2001). *Then & now: Changes in child care staffing, 1994–2000*. Technical Report. ERIC.

### Publisher's Note

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

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---