Relationship between teaching experience and teacher effectiveness: implications for policy decisions

Jeff Irvine
Brock University

ABSTRACT

This study examined claims that teachers' years of teaching experience correlate to teachers' effectiveness. The assumed experience–effectiveness relationship was used to support the Government of Ontario, Canada's policy decisions concerning teacher hiring practices. This study critically examined sources cited in the policy report and reviewed other research on teacher effectiveness. Findings indicate that the relationship between total years of experience and teacher effectiveness, as measured by student achievement gains, is complex, nuanced, and nonlinear. The conclusion is that decisions based on assumptions that the relationship between experience and effectiveness is direct and linear are simplistic and lead to less than optimal policy.

Keywords: teacher effectiveness, teacher experience, VAM

We all want our children to have the best, most effective teachers. Unfortunately, there is little agreement on how to measure or what constitutes effectiveness. It is tempting to equate effectiveness with years of teaching experience—that is, to assume that more experienced teachers are more effective. This paper explores the relationship between experience and effectiveness, and how this relationship was used in part to support an educational policy decision. Provincial and/or state educational jurisdictions frequently make policy decisions based on research evidence; however, such research is not summarized directly by the educational authority but rather is filtered through external agencies such as universities or private research facilities. In this case, the research evidence that supported the decision was not very strong, and thus the policy may not be optimal.

Context of the Study

"All other things being equal, teachers with more experience are better teachers" (Directions Evidence and Policy Research Group [DEPRG], 2014, p. 25). The latter statement is from the final report commissioned by the Ontario Ministry of Education on Regulation 274 regarding teacher hiring practices. This study considered the veracity of DEPRG's claim by first deconstructing the sources cited in its report and then examining other research literature on the relationship between teacher experience and student achievement.

Ontario Regulation 274

The province of Ontario has a surfeit of qualified teachers, with an estimated 30,000 qualified educators currently eligible "for jobs that don't exist" (Wente, 2013, para. 4). Prior to 2012, it was not uncommon to have hundreds of applicants for any permanent teaching position advertised, with the majority of such applicants being drawn from the substitute teacher roster in the district in which the position was posted. The majority of candidates in the substitute teacher roster in turn were qualified novice teachers who were graduates of a faculty of education. Additionally, some qualified teachers had experience in another district, another jurisdiction, or sometimes in another country. Substitute teachers replaced those who were absent from class for a short time due to reasons such as illness, professional development, or conference attendance. A substitute teacher usually was provided with complete lesson plans, and was not responsible for long-term planning, assessment, or reporting.

In addition, substitute teachers could apply for long-term occasional (LTO) positions, which involved assignments in the same class for 15 or more consecutive days. These situations occurred due to regular full-time teachers' maternity leaves, long-term illness, or personal leaves for other reasons. LTO teachers typically assumed the full duties of the teachers they replaced, including lesson planning, assessment, and reporting.

The Ontario government expressed concern about the inherent subjectivity of the interview process for permanent positions, and the possibility of favouritism or even nepotism. To address this concern, the government enacted Regulation 274, which standardized teacher hiring practices and established an LTO-teacher list. To qualify for the list, a substitute teacher had to have completed at least one LTO assignment, with a duration of 4 months or more, and teachers on the list were then ranked by seniority. The regulation required school administrators hiring for a permanent position to select from the top five teachers on the LTO list in terms of seniority, subject to those teachers being qualified for the open position (Ontario Ministry of

Education, 2012). School administrators expressed a number of concerns regarding this regulation. Chief among these concerns was administrators' preferred use of an interview process to determine how well the candidate *fit* the school and its student population, which thus took into account factors such as pedagogical style, extracurricular contributions, and potential role models for students in a given school.

In 2014, Ontario commissioned the DEPRG report on Regulation 274, which included the statement corresponding to teacher experience and teacher quality noted at the outset of this paper. Ultimately, the popular press characterized Regulation 274 as an instrument that required teachers to be hired based on seniority rather than on merit. The present study therefore sought to examine the relationship between teacher experience and teacher effectiveness and to investigate the impact of Ontario Regulation 274.

Methodology and Methods

This study utilized constructivist grounded theory (Charmaz, 2014), which makes no a priori assumptions about whether a relationship exists between teacher effectiveness and teacher experience, nor the form such a relationship might take. Rather, the grounded theory methodology (known as the constant comparison method) employs an iterative approach of data collection, data analysis, and additional data collection based on this analysis, until data saturation is reached, and no new insights or new properties are generated by further data collection. Constructivist grounded theory employs multi-level coding, proceeding from initial open coding to more focused coding once themes have emerged, to axial coding, which relates coding categories to subcategories, and finally to theoretical coding that links categories to produce a hypothesis or theory (Noerager Stern & Porr, 2011). Noerager Stern and Porr (2011) describe grounded theory as a method of seeking an "inductive generalized pathway as opposed to a deductive verificational pathway" (p. 39). Constructivist grounded theory is a particularly appropriate tool for textual analysis (Charmaz, 2014) and well-suited for policy analysis (Richards & Farrokhnia, 2016). Data analysis is continued until no new information or relationships are identified (Sutcliffe, 2016).

Constructivist grounded theory resulted in several additional research questions in addition to the principal research question about the relationship of teacher experience to teacher effectiveness.

Research Questions

The study's primary goal to identify the relationship between teacher experience and teacher effectiveness was guided by the following research questions:

1. What are the attributes of effective teaching?
2. How is effective teaching measured?
3. What are typical effect sizes for effective teaching with respect to student achievement?
4. What has been the impact of Ontario Regulation 274 on student achievement, and what alternatives should be considered for this regulation?

There is little doubt that student achievement is impacted by teacher quality, sometimes referred to as teacher effectiveness (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2008a, 2008b; Boyd, Lankford, Loeb, & Wyckoff, 2010; Clotfelter, Ladd, & Vigdor, 2007a; Fantilli & McDougall, 2009; Goldhaber & Anthony, 2007; Harris & Sass, 2009; True, Butler, & Sefton,

2011). Clearly, education systems want to engage and retain the most effective teachers; however, questions remain regarding what constitutes effective teaching, how to measure effectiveness, and whether effectiveness and experience are synonymous.

It is tempting to link observable teacher characteristics such as certification, years of experience, and advanced academic degrees to teacher effectiveness and student achievement. Such characteristics are perceived as more objective and more easily measurable, and therefore are used in many jurisdictions to identify teacher salary levels. Unfortunately, the research typically does not support the link between these characteristics and student achievement gains.

Characteristics of Effective Teachers

The metric for teacher effectiveness is student achievement (Doherty & Jacobs, 2013), but there is little agreement on what constitutes achievement. The most common definition of achievement corresponds to increases in standardized test scores. This teacher value-added measures (VAM) approach, discussed below, is a very narrow definition of achievement, and can be applied only to about 20% of teachers in the United States (Staiger & Rockoff, 2010). A number of other dimensions may be considered, such as student motivation, engagement, persistence, student attitudes, student self-efficacy, and student self-concept. These dimensions are typically treated as mediator variables (motivation, engagement, persistence) with respect to student achievement or moderator variables (attitude, self-efficacy, self-concept) with respect to student achievement. Still, it can be argued that student increases in these dimensions are also outcome variables that represent student achievement.

Because observable teacher characteristics such as certification, advanced degrees, and years of experience frequently impact teacher salary conditions, many of the VAM studies in these areas are undertaken by economists as well as educational researchers. This often results in the use of education production–function models (e.g., see Rockoff, 2004) that focus on a process–product formulation, viewing student achievement as the product and teaching as the process. This is a somewhat narrow view of student achievement, often omitting numerous important student outcomes such as informed citizenship, leadership development, cooperation, and decision-making. Cohen and Goldhaber (2016) point out that

> There are multiple dimensions by which a teacher might be deemed "effective." Teachers support students in myriad ways in service of multiple outcomes, including social and emotional outcomes. Even if we focus on the narrower construct of teacher effectiveness as supporting student learning on academic outcomes, value-added estimates culled from a single test may only represent a portion of the broader construct of interest. (p. 380)

Palardy and Rumberger (2008) examined three dimensions of teacher effectiveness—instructional practices, teacher attributes (such as self-efficacy, attitude, enthusiasm), and teacher background characteristics (certification, advanced degrees, years of experience)—in relation to learning among students in Grades 4 to 8 in New York City using hierarchical linear modeling (HLM) that separates the effects of various elements, such as students, teachers, schools, and school districts. With respect to teacher attributes, Palardy and Rumberger found that instructional practices had the most direct impact on student achievement, followed by indirect effects of teacher attributes. Teacher background characteristics such as total years of experience and certification were found to have no effect on student achievement.

A number of studies have connected pedagogical content knowledge as well as subject matter knowledge to teacher effectiveness. König and Kramer (2016) identify pedagogical

content knowledge as "one of the central cognitive components of teacher competence" (p. 142). This attitude was echoed by Ingvarson and Rowley (2017), whose analysis of teacher education programs in 17 countries found that pedagogical content knowledge was an important indicator for identifying potentially effective new teacher candidates. Numerous studies identify both pedagogical content knowledge and subject matter knowledge as important for effectiveness in teaching mathematics (e.g., Blazar & Pollard, 2017; Fagginger Auer, Hickendorff, Van Putten, Beguin, & Heiser, 2016; Krauss, Baumer, & Blum, 2008).

Another dimension of effective teaching is teacher enthusiasm and motivation of students to succeed. Motivation is typically treated as a mediator variable with respect to student achievement; however, Frenzel, Goetz, Ludke, Pekrun, and Sutton's (2009) study of emotional transmission between teachers and students found that "Classrooms, which are characterized by enjoyment of teaching *and* learning likely provide optimal grounds for overcoming obstacles and promoting positive development and achievement" (p. 712). Frenzel et al. thus recognized that achievement is only one aspect of student development, and that positive attitudes and student growth in such areas as self-efficacy and persistence are also important outcomes. This stance was affirmed by Hill, Blazar, and Lynch (2015) in their study of teacher personal and institutional predictors of high-quality instruction. Kunter, Tsai, et al. (2008) in turn found that students' and teachers' perceptions of teacher enthusiasm in mathematics classes had a significant effect on instruction and hence on student achievement; key elements of teacher effectiveness were found to be teacher enthusiasm, cognitive challenging tasks, learning support, and classroom management. In a later study, Kunter, Klusmann, et al. (2013) reaffirmed the elements listed above, and added pedagogical content knowledge as an important element in effective teaching. Kunter, Klusmann, et al. also found that student achievement in mathematics was improved when teachers had constructivist beliefs in the teaching and learning of mathematics.

Several mathematics education studies have identified the term *ambitious teaching* to describe effective teaching in mathematics. Lampert et al. (2013) describe the goal of ambitious teaching as preparing teachers to "do more socially and intellectually ambitious [teaching] than the current norm" (p. 226). Anthony and Hunter (2013) summarize ambitious mathematics teaching as teaching "in which conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition are intertwined in mathematical practice and learning" (p. 699). Again it is noteworthy that "productive disposition" is identified as an outcome variable. Ambitious teaching was found to be difficult to inculcate in new teachers due to the impact of their prior experiences as students, typically in transmission-oriented classrooms (Hiebert, Morris, Berk, & Jansen, 2007).

Hattie (2010) labelled effective teachers as *expert teachers*. While Hattie found no evidence supporting a link between teacher experience and student achievement, he determined that expert teachers have very significant impacts. Hattie identifies five dimensions that set apart expert teachers; they can: identify essential representations of their subject, guide learning through classroom interactions, monitor learning and provide feedback, attend to student affective attributes, and influence student outcomes. Hattie also distinguishes 16 subdimensions within the five macro dimensions, and he elaborates on each of them. Overall, expert teachers differ from experienced teachers in a number of important ways. For example, Hattie found that students' work in experienced teachers' classrooms was focused on surface learning 72% of the time, and only 28% on deep learning. In contrast, the opposite was seen in student work from

expert teachers' classrooms, whereby the work focused on deep learning 72% of the time and only 28% on surface learning.

There is some danger in using the term *expert* with respect to teaching. For example, Palmer, Stough, Berdenske, and Gonzales's (2001) investigation of the characteristics of experts in various fields—including athletes, musicians, and chess masters—and attempt to link these characteristics to teachers reiterated the 10,000 hours of deliberate practice requirement for becoming an expert; however, all the professions investigated in Palmer et al.'s study possessed an internal lens for defining expert. For example, the study enumerated characteristics necessary for a person to become an expert musician—not to teach others to become one. As a result, this lead to a set of expert criteria for teachers that did not reference students or student achievement at all.

How Is Teacher Effectiveness Measured?

Three benchmarks are typically used to measure teacher effectiveness: VAM, classroom observations, and surveys (student responses, peer responses, administrator responses).

Value-Added Measures

Much of the more current research pertaining to the teaching experience–effectiveness relationship uses VAM that examine student achievement *gains* rather than absolute levels of student achievement (sometimes called *status* models). Such research focuses on increases in student achievement in a particular time frame (usually an academic year, but sometimes over several years) and relates these gains to teacher or school effects. In this way, the studies control for the effects of much of the student gains attributable to external factors, such as home effects or students' prior knowledge. VAM are very common in U.S.-based studies on education, with student scores on standardized tests used for the before-and-after measures of achievement. It is estimated that a student in the U.S. public education system will experience 112 standardized tests between pre-Kindergarten and the end of Grade 12 (Casserly, 2015). Standardized tests are an imperfect method of measuring a number of important student achievement categories, such as problem solving, but have some benefits in being relatively objective measures (Stronge, 2010).

Additional concerns with VAM include the selection effect through which less effective teachers leave the profession, resulting in an increase in the remaining teachers' overall average effectiveness (Boyd et al., 2008a, 2008b; Huang & Moon, 2009; Jacob, 2012). This also introduces sample bias because the remaining teacher pool has been modified (Jacob, 2012). In addition, effective teachers may be assigned classes with high-performing students as a reward or, conversely, classes of underperforming students in an effort to raise overall student achievement; in either case, teacher assignment is not random (Huang & Moon, 2009; Jacob, 2012). All of these issues impact the findings of VAM studies.

The American Educational Research Association (AERA, 2015) recently issued a statement on the use of VAM, particularly as they pertain to teacher evaluation, training, retention, and dismissal, as well as the use of VAM in evaluating teacher preparation programs. The AERA statement identifies eight conditions that must be addressed for VAM models to be valid and reliable measures of teacher evaluation. Many studies do not conform to these conditions and AERA warns against extrapolating results based on studies lacking one or more of the conditions. Darling Hammond (2015) identified a number of negatives with respect to

using VAM, including the differential effect on non-random student assignment, in which some teachers are more effective with some students or types of students than others, and identified an equity issue, since VAM are frequently only applied to teachers of mathematics and reading, thus excluding over 70% of teachers. Darling Hammond also found that VAM cannot disentangle other influences on student achievement from teacher effects, and that measures of effective teaching using VAM are highly unstable over time. Broatch and Lohr (2012) argue that VAM are too short term to evaluate a complex activity such as teaching, and that a longer-term view of teacher effectiveness is necessary.

In sum, VAM are only applicable to a relatively small sample of teachers for whom sufficient data are available. For example, Atteberry, Loeb, and Wyckoff (2015) found that only about 20% of the teachers in their sample had sufficient data to employ VAM. Atteberry et al. found that VAM for novice teachers are less predictive of future effectiveness than VAM for more experienced teachers.

The *Measures of Effective Teaching Project* (MET Project; Bill & Melinda Gates Foundation, 2010, 2013) used VAM as well as three supplementary student assessments in addition to state standardized tests, and also evaluated teacher pedagogical content knowledge. To obtain a richer data set on teacher effectiveness, the MET Project used student questionnaires about students' perceptions of teacher effectiveness. Although the project determined that VAM for early-career teachers is a strong predictor of future performance, Staiger and Rockoff (2010) found that VAM for early-career teachers showed small effect sizes and high levels of "noise" in the data, with variances in excess of double the effect sizes. This makes efficient predictions of future effectiveness difficult.

Ruzek, Domina, Conley, Duncan, and Karabenick (2015) attempted to apply VAM to non-traditional student achievement measures such as changes in student motivation. In their study of 2,864 middle-school (Grade 7) students in California, Ruzek et al. applied VAM measures to teacher-associated changes in students' achievement goal orientation in mathematics. Pre- and post measures of students' three-goal orientations (mastery, performance-approach, and performance-avoidance) found that students in classes with more effective teachers showed gains in both mastery and performance-approach orientation as well as gains in achievement. Ruzek et al.'s study emphasized the differential teacher effects on students and identified student non-cognitive factors such as motivation as an important consideration for policy makers.

Students in Ontario, Canada face very few standardized tests. Criterion-referenced assessments are administered in reading, writing, and mathematics in Grade 3 and Grade 6; in mathematics in Grade 9; and in literacy in Grade 10. Only the latter test is high stakes, as passing it is a graduation requirement for secondary school. All the assessments cover a range of grades, and thus a range of teachers, and therefore are not useful for assessing teacher effectiveness using VAM.

Typical Effect Sizes for Effective Teaching With Respect to Student Achievement

The effect size of teacher experience, even in the first 3 years of teaching, is disappointingly small. For example, in Grades 4 and 5, the effect size was 0.06 standard deviations of student achievement, and even smaller in middle school (Grades 6 to 8) at 0.04 standard deviations (Boyd et al., 2008a; King Rice, 2010). There was some differentiation in effect size based on subject, ranging from 0.06 standard deviations in mathematics to 0.03

standard deviations in reading (King Rice, 2010). Atteberry et al. (2015) segmented novice teachers of Grades 4 and 5 by quintiles based on VAM (a posteriori) and found that even teachers in the fifth quintile over the first 5 years of teaching had effect sizes of only 0.15 in mathematics and 0.08 in language arts. Teachers in the lower three quintiles continued to have small negative effect sizes over the entire 5-year period. Staiger and Rockoff (2010) found effect sizes of 0.09 to 0.16, but these estimates had low reliability, and the "noise" in the data was more than double the effect sizes found.

Relationship Between Teacher Experience and Teacher Effectiveness

As noted earlier, the DEPRG (2014) report claims that "All other things being equal, teachers with more experience are better teachers" (p. 25). This section discusses a number of sources cited in the DEPRG report, followed by other research relating teacher experience and teacher effectiveness.

Sources Cited in *Ontario Regulation 274—Final Report*

Five sources cited in the DEPRG (2014) report relate to teacher experience and student achievement. No rationale is provided for the selection of these specific studies. The first is Biniaminov and Glasman's (1983) work that examines links between teacher experience and Grade 12 graduation rates in Israel (for the 1975-76 school year). The authors test a number of models, only some of which indicate a relationship between teacher experience and graduation rates, and only for years of teacher experience *in the same school*. There are a number of problems with the use of Biniaminov and Glasman's study as supporting evidence to suggest that total years of teaching experience relates to teacher quality, and hence to student achievement.

First, the data in Biniaminov and Glasman's (1983) study are 40 years old, and hence DEPRG instead could have cited many more current studies. Second, the study uses the school rather than individual teachers as the unit of analysis, hence the teacher-experience variable is predicated on average number of years of experience in the same school rather than on teachers' total years of experience. The study also uses the proportion of Grade 12 students passing certification exams (i.e., gross student achievement rather than achievement gains), which is the more common measure of teacher effectiveness. Using absolute achievement makes it more difficult to disentangle teacher effects from other factors, such as student, home, and socioeconomic effects. In their introduction, Biniaminov and Glasman state that "Agreement has been only moderate that school variables, in fact, influence achievement" (p. 251). While such a statement may have been true when the paper was published in 1983, it certainly is less accurate in today's research environment. Blazar (2015) cites multiple sources in support of his claim that "Over the past decade, research has confirmed that teachers have substantial impacts on their students' academic and life-long success" (p. 16). Hattie (2003), for example, estimates that teacher effects account for 30% of the variance in student achievement scores, and the school an additional 5% to 10%; thus, school effects are second only to student effects (50%) in accounting for the variance in student achievement. Seebruck (2015) found that students who were taught by highly effective teachers for 3 years showed gains of 35 and 50 percentile points in reading and mathematics, respectively.

More problematic still are two a priori statements in Biniaminov and Glasman's (1983) paper. The first states that one of the study's independent variables is "teachers' experience in

the same (current) school which, in part, reflect teacher quality" (p. 252), which is an assumption made without any supporting evidence prior to the commencement of the study. The second a priori statement—"Because teachers with longer experience are better acquainted with the abilities of specific student populations and with requirements of government examinations, they are better equipped to facilitate students' academic success on these examinations" (p. 258)—while perhaps true, would be better suited to the paper's discussion section, after presentation of the study's results. Establishing this statement a priori creates another assumption which, like the aforementioned one, may have biased the researchers' analysis. In summary, Biniaminov and Glasman's study provides very weak supporting evidence and is not a rigorous source in the DEPRG report.

The second source, Huang and Moon's (2009) multilevel analysis of teacher characteristics and student achievement, is a more recent and a much stronger paper. It examined possible linkages in a data set consisting of 1,544 Grade 2 students, 154 teachers, and 53 schools in a Mid-Atlantic state of the United States. Huang and Moon found that total years of experience was not significantly associated with student achievement, although years of experience *in the same grade* was positively associated in two cases. In their literature review, Huang and Moon note that while 3% of the examined studies actually indicated a negative association between teachers' years of experience and student achievement, and 30% indicated a positive association (usually for only the first few years of teaching), the majority (67%) showed no correlation between teachers' years of experience and student achievement. Huang and Moon found little evidence of teacher improvement after the third year of teaching and that total teaching experience was not significant in any of the cases they investigated.

The third source, Jepsen and Rivkin's (2009)—cited erroneously as 2007 in the DEPRG report—study of class-size reduction and student achievement examined a 1996 California mandate reducing class sizes in the early primary grades to 20 students per class, an initiative that cost $1 billion and necessitated the hiring of 25,000 additional teachers. Many of these new teachers had no previous teaching experience, and many were uncertified as well. Jepsen and Rivkin found that the class-size reductions' impact on student achievement was moderated by the influx of inexperienced teachers; however, they also noted that "observed teacher characteristics like experience and certification explain very little of the variation in teacher effectiveness" (p. 241).

The fourth source, Mulholland and Berliner's (1992) *Teacher Experience and the Estimation of Student Achievement*, is not at all on point. The study examines differences between pre-service and experienced teachers' ability to predict students' final marks. Not surprisingly, experienced teachers who have a greater understanding of their students were better able to predict student achievement. However, the paper does not at all touch on the relationship between teachers' years of experience and student achievement.

The final source, Nye, Konstantopoulos, and Hedges's (2004) study of teacher effects, examined teachers' years of experience and student achievement in Grades 1 to 3. Nye et al. found limited associations, again in the first 3 years of teaching, but also that teacher experience in any year explained very little and never more than 5% of the variance in student achievement. Nye et al. postulate a nonlinear relationship between teacher experience and student achievement, which is consistent with Huang and Moon (2009) who found a parabolic relationship for teachers who employed flexible groupings. Huang and Moon found that less experienced teachers were more likely to employ flexible groupings and that the frequency of using groups increased through the first several years of teaching but then peaked and began to

decline for teachers with longer time in the classroom. Nye et al. conclude that "there are substantial differences among teachers in the ability to produce achievement gains in their students" (p. 253), but teachers' years of experience were generally not a significant factor.

In sum, the sources cited in the DEPRG (2014) report provide little support for the statement that "all other things being equal, teachers with more experience are better teachers" (p. 25). Indeed, one source is completely off point while another is significantly out of date and indicates weak scholarship and confounding a priori assumptions. The other three sources all contain statements indicating that total years of teaching experience are not significantly related to student achievement. Thus, the DEPRG report's claim that "all teachers with more experience are better teachers" cannot be considered valid based on the evidence cited.

Other Studies on the Teacher–Experience and Student–Achievement Relationship

Harris and Sass (2009) found that teacher effectiveness is influenced significantly by teacher personality factors, especially caring, intelligence, subject matter knowledge, and teaching skills. Harris and Sass compared teacher effectiveness results from VAM used in Florida elementary, middle, and high schools as well as from principal interviews based on teacher observations. They found that observation of teachers by principals is a superior method of assessing teacher effectiveness compared to VAM; however, there are issues associated with observational techniques, including subjectivity, convenience, and cost (Cohen & Goldhaber, 2016). In addition, when teacher evaluation ratings are binary (satisfactory/unsatisfactory), Weisberg, Sexton, Mulhern, and Keeling (2009) found that 99% of teachers are rated "satisfactory." Weisberg et al. noted that even when evaluation scales are broader and have multiple rating categories, 94% of teachers are given one of the top-two ratings. A second examination of teacher evaluation systems by Kraft and Gilmour (2017) found that although most states had modified their teacher evaluation systems, the percentage of teachers rated satisfactory remained at 99%. This contrasts sharply with Staiger and Rockoff's (2010) research, which used Monte Carlo methods showing that in order to optimize average teacher effectiveness, as measured by VAM, 80% of new teachers need to be dismissed after their first year. While based on simulations and not actual teacher effectiveness data, Staiger and Rockoff demonstrated the disproportionate effect of novice teachers when VAM are employed as the teacher effectiveness measure.

Many studies indicate that teacher experience influences teacher effectiveness mainly in the first few years of a teacher's career (e.g., Clotfelter et al., 2007a, 2007b). Boyd et al. (2010) claim that improvement occurs over the first 4 or 5 years, but point out that this is an average, and some less experienced teachers improve much more quickly than others. Many of the studies that found student achievement was influenced by teachers' years of experience identified these effects only in the first 3 years of a teacher's career: "The average teacher is at his or her worst during the first year in the classroom, gets better in the second year, a little better in the third year, and then never gets any better after that" (Jacob, 2012, p. 3).

Researchers found that the relationship between student achievement and teachers' years of experience was highly nonlinear, peaking quickly after the first few years of teaching: "Strong evidence suggests that teacher effectiveness spikes sharply after the first few years in the profession" (Fantilli & McDougall, 2009, p. 814; see also Hanushek, Kain, O'Brien, & Rivkin, 2005; Huang & Moon, 2009; Jacob, 2012; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004). Rockoff (2004) actually found that having more than 2 years of teacher experience was

detrimental to student achievement in math computation and that years of experience had no impact on math concept attainment. This deterioration in teacher effectiveness also was noted in Clotfelter et al.'s (2007b) study of high school teachers, and in King Rice's (2010) study that similarly found significantly diminishing teacher effectiveness in high schools, particularly in mathematics teachers. Huang and Moon (2009) found a positive relationship between student achievement and teachers' years of experience in the same grade. Many of the studies are confounded by other factors, such as early leaving by ineffective teachers with less experience, and attrition of teachers with greater experience.

King Rice's (2010) examination of teacher experience's impact on policy concluded that the "more is better" assumption is simplistic and that the relationship between teacher experience and teacher effectiveness is more complex and affected by multiple factors. King Rice states that "The impact of experience is strongest during the first few years of teaching; after that, marginal returns diminish" (p. 1). King Rice reiterates the generally negligible values of the impact of experience during the first few years of teaching (0.04 to 0.06 standard deviations) and the confounding effects of differential attrition in teachers, and she also notes that the results are averages; thus, individual teachers' effectiveness may vary dramatically from these averages. Further, the magnitude of teacher experience effects differs depending on a teacher's level of education and subject area.

King Rice (2010) also found a significantly uneven distribution of both inexperienced and ineffective teachers, with a preponderance of the latter group found in socioeconomically disadvantaged schools. King Rice determined that even experienced teachers' effectiveness declined when they were placed in high poverty schools. King Rice speculates that this may be because these teachers are ill-equipped to deal with the challenges in high poverty schools, and/or that such challenges may result in teacher burnout.

Hattie (2003) parses the variance in student achievement into the following categories: students, 50%; home, 5% to 10%, mainly via expectations and encouragement; schools, 5% to 10%; principals (included in the schools category); peers, 5% to 10%; and teachers, 30%. Thus, teachers have the greatest controllable in-school impact on student achievement. As noted earlier, Hattie makes the important distinction between experienced teachers and expert teachers.

Conclusion

There is little doubt that teacher quality impacts student achievement (Goldhaber & Anthony, 2007; Hattie, 2003; Jacob, 2012; Kane, Rockoff, & Staiger, 2008; Palardy & Rumberger, 2008; Stronge, 2010); however, identifying the attributes of effective teachers and determining how to measure such attributes is more challenging. While it is tempting to assess effective teaching through easily observable attributes such as years of experience or academic degrees, these are "at best weak predictors of a teacher's contribution to student achievement" (Harris & Sass, 2009, p. 1; see also Winters, 2011). The linking of teacher effectiveness to years of experience is very tempting, based on the craftwork concept that novice teachers apprentice at the hands of a master teacher. However, this analogy does not fit the reality of teaching, which is often a solitary profession in which most learning by the teacher is based on reflective practice—both reflection-on-action and reflection-in-action (Schön, 1983). Gains in teacher effectiveness during the first few years are typically in classroom management (König & Kramer, 2016) and effect sizes are small. While new teacher induction and mentoring programs have been shown to be very useful (Ingersoll & Strong, 2011), there is little uniformity in these mentoring programs

and often little monitoring of results. For example, in Ontario's new teacher induction program (NTIP, Ontario Ministry of Education, 2006), implementation rests with voluntary mentors who have no additional training and are sometimes reluctant to take on extra responsibilities with no concomitant recognition or release time.

As discussed in this paper, the relationship between teacher experience and teacher effectiveness is problematic. While many studies indicate teachers' effectiveness improves in "the first few years" (typically the first 3 years), average effect sizes are small and do little to indicate the quality of a specific teacher (King Rice, 2010). Furthermore, several studies refute the "more experience is better" mantra (Boyd et al., 2010; Clotfelter et al., 2007a; Fantilli & McDougall, 2009; King Rice, 2010). In their extensive literature review, Huang and Moon (2009) found that while 3% of the examined studies indicated a negative association between teachers' years of experience and student achievement, and 30% indicated a positive association (usually for only the first few years of teaching), the majority (67%) showed no correlation between teachers' years of experience and student achievement. Thus, using teacher experience as an indicator of teacher effectiveness appears to be a poor choice.

Teacher effectiveness is measured by student achievement. Other affective aspects such as student motivation, attitude, and engagement typically have been shown to be mediator variables or moderator variables with respect to student achievement (Baron & Kenny, 1986). Measuring teacher effectiveness is problematic; VAM have a number of issues, and frequently insufficient data is available (Atteberry et al., 2015). Other measures of teacher effectiveness such as principals' observations or rubrics have additional issues such as cost and subjectivity. Wilson, Hallam, Pechone, and Moss (2014) found that including teacher portfolios as part of the evaluation process increased the reliability of identifying effective teachers; however, portfolio assessment is difficult to implement for novice teachers who may only have practice teaching evaluations and limited sample artifacts available. In addition, portfolio assessment may be subjective, since it is unlikely that uniform rubrics are available to all administrators. As part of the *Measures of Effective Teaching Project (MET)*, Kane, McCaffrey, Miller, and Staiger (2013) identified three measures of teacher effectiveness: VAM, student surveys, and classroom observations. Kane et al. validated these measures using random assignment, but warned that their conclusions were still subject to prediction errors and restricted to random assignment within a specific school, thus limiting the generalizability of their results. An attempt by Mihaly, McCaffrey, Staiger, and Lockwood (2013) to construct a composite estimator of effective teaching noted the relative lack of objective data. Mihaly et al. also noted that many states in the United States now require additional indicators of teacher effectiveness such as classroom observations, teacher reflections, student surveys, and identified student learning objectives.

The costs of making an error in new teacher hiring are large. Without an effective teacher, student costs include suboptimal achievement, motivational declines, possible social and peer stigmatization, and negative emotional impacts. More broadly, the community and school district are impacted by financial and training or support costs and opportunity costs associated with not hiring a better candidate. The less-effective teacher also faces financial costs and emotional costs associated with failure (one-third of teachers leave the profession within the first 5 years; Darling-Hammond, 2003). This trend has important implications for student achievement. "Probably the most important thing a school administrator at the school or district level can do to improve student achievement is to attract, retain, and support the continued learning of well-prepared and committed teacher" (Darling-Hammond, 2003, p. 2).

Ontario Regulation 274

There are grave implications to utilizing any sorting or hiring system that is not research-based. For example, Florida recently created a bonus system called "Best and Brightest" that rewards teachers based on their own standardized test scores from their high school years. This system has been implemented despite absolutely no research evidence that teachers' high school test scores have any correlation to teacher effectiveness or student achievement (Morgan, 2017).

With respect to research question 4, there are no data on the impact of Ontario Regulation 274 on student achievement. Ontario utilizes some criterion-referenced standardized tests administered by the Education Quality and Accountability Office (EQAO), but these tests are reported by schools and are not linked to individual teachers. Thus, utilizing VAM to investigate teacher effectiveness is not possible. There is anecdotal evidence that school principals are dissatisfied with the results of Regulation 274 in that they felt forced to hire teachers whom they would not have otherwise considered. Principals have expressed frustration that they are held responsible for their schools' EQAO results, but they have severely limited autonomy in hiring teachers. Some of these teachers have had classroom management issues, and others have come from jurisdictions that emphasize rote memory and lectures, which is not compatible with Ontario's student-centred and social constructivist paradigm. A survey of 2,050 principals by the organization People for Merit-Based Teacher Hiring (Levy, 2015) found that 96% of principals in Ontario felt that seniority-based hiring is not in the best interests of students and 89% felt that they were prevented by Regulation 274 from interviewing teacher applicants who best matched the needs of their student community; moreover, 84% claimed that the current teacher hiring process hinders student learning, while only 1% believe it has improved learning.

Further, the majority of studies relating teacher experience and teacher effectiveness focus on teachers who have the same class or classes for an entire year. With Ontario Regulation 274, teachers on the LTO list may have as few as 4 months of experience, and any research conclusions may not apply. Because research shows that teachers form their permanent teaching styles during their first year of teaching (Weiss, 1999), there is a concern that some of these newly hired teachers may have inadequate foundations in pedagogy and particularly in student assessment practices as outlined by the Ontario Ministry of Education.

Consideration must also be given to Hattie's (2003) distinction between experienced and expert teachers. From the 16 subdimensions that differentiate expert teachers from experienced teachers, Hattie identifies a *minimal subset* of three dimensions that allow the classification of teachers as expert versus experienced with 80% accuracy. This minimal subset consists of deep representations of teaching and learning; providing appropriately challenging tasks and goals; and monitoring student problems, assessing their level of understanding and progress, and providing relevant, useful feedback—none of which are directly related to experience, nor can be assessed based on number of years of teaching. This is hugely problematic, given Regulation 274's reliance on mere seniority on a list as the requirement for permanent teaching positions. In short, this is a simplistic solution to a complex problem. While it is possible that good policy decisions may result despite weak research evidence support, this is clearly not an optimal path to decision making. It is possible that this regulation addressed a political goal, providing the appearance of equity in teacher hiring. Nonetheless, such a goal does nothing to enhance teacher effectiveness and student achievement.

Staiger and Rockoff (2010) point out that

> The current system, which focuses on credentials at the time of hire and grants tenure as a matter of course, is at odds with decades of evidence on teacher effectiveness. Instead, teacher recruitment and retention policies should focus on improving our methods of teacher evaluation and use admittedly imperfect measures of teacher effectiveness to identify and retain only the best teachers early in their teaching careers. (p. 115)

Ontario Regulation 274 completely ignores issues of teacher effectiveness and settles for a simplistic measure—namely, seniority on an LTO teacher list. This is an imperfect measure even of teacher experience and has little or no relationship to hiring effective teachers.

Other options to Regulation 274 should have been considered. For example, standardized interview questions with a rating scale (1 to 5) for respondents' answers would reduce the inherent subjectivity of the interview process. Alternatively, interviews could adopt a standardized provincial interview rubric that addressed key features of educational policy in Ontario. The rubric could identify such areas as student-centred instruction, Ontario assessment policies, the social constructivist nature of learning, effective questioning, and other aspects of teaching policy in the province.

Another potential alternative is to utilize videotaped teacher lessons. Kersting, Givvin, Thompson, Santagata, and Stigler (2012) found that observing videotapes of teachers in action allowed trained raters to impute levels of teacher content knowledge and content knowledge for teaching based on teacher interactions with students. Santagata and Guarino (2011) used videotapes of classroom lessons to quantify teacher reflective practices, including reflection-in-action and reflection-on-action (Schön, 1983). For teacher hiring, the top five candidates for a permanent teaching position could be required to submit a videotape of a sample lesson, which could then be assessed using a provincially generated rubric or rating scale. Because classroom practices are crucial to student achievement, this method should increase the probability that the best teacher is hired, using student achievement as the metric.

Finally, Hattie's (2003) finding that students' work in experienced teachers' classrooms was focused on surface learning 72% of the time, and only 28% on deep learning—while the opposite was found in student work from expert teachers' classrooms, whereby the work focused on deep learning 72% of the time and only 28% on surface learning—could provide a basis for evaluating practising teachers to identify teacher effectiveness. This is an objective e method for identifying effectiveness that would replace or supplement the current satisfactory/unsatisfactory system, which has been shown to be virtually non-evaluative, as 99% of teachers are rated satisfactory (Kraft & Gilmour, 2017).

Stronge's (2010) summary of the relationship between teaching experience and student achievement emphasizes that 1 to 3 years of experience generally has a positive effect, as does teaching experience in the same grade, but research on the relationship between years of experience beyond 3 and student achievement remains inconclusive. However, Jacob's (2012) broad study of Florida elementary student data found that "Upwards of 97% of what makes one teacher more effective than another is unrelated to factors such as the number of years the teacher has been teaching and the credentials that the teacher has earned" (p. 3). As such, using seniority as a proxy for teaching experience and hence as a criterion for hiring permanent teaching positions is ill-advised and should be reconsidered.

In a recent press release for the provincial government's Education and Equity Action Plan, Ontario's Minister of Education (2017) stated that

> The plan says staff needs to better reflect the diversity of the student body, adding that senior administrators will be encouraged to keep equity in mind when hiring and

> promoting staff and that such efforts will be included in their performance appraisals. ("Ontario Spending Millions," 2017, para. 7)

However, the minister gave no indication that Regulation 274 would be modified, so how these conflicting objectives can be met is unclear.

References

American Educational Research Association. (2015). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation models. *Educational Researcher*, *44*(8), 448-452. doi:10.3102/0013189X15618385

Anthony, G., & Hunter, R. (2013). Learning the work of ambitious mathematics teaching. In V. Steinle, L. Ball, & C. Bardini (Eds.), *Mathematics education: yesterday, today, and tomorrow. Proceedings of the 36th annual conference of the Mathematics Education Research Group of Australasia* (pp. 699-702). Melbourne, Australia: Mathematics Education Research Group of Australasia.

Atteberry, A., Loeb, S., & Wyckoff, J. (2015). Do first impressions matter? Predicting early career teacher effectiveness. *AERA Open*, *1*(4), 1-23. doi:10.1177/2332858415607834.

Baron, R., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173-1182. doi:10.1037/0022-3514.51.6.1173

Bill & Melinda Gates Foundation. (2010). *Learning about teaching: Initial findings from the Measures of Effective Teaching Project* (MET Project research paper). Retrieved from https://docs.gatesfoundation.org/Documents/preliminary-findings-research-paper.pdf

Bill & Melinda Gates Foundation. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study* (MET Project research paper). Retrieved from https://www.edweek.org/media/17teach-met1.pdf

Biniaminov, I., & Glasman, N. (1983). School determinants of student achievement in secondary education. *American Educational Research Journal*, *20*(2), 251-268. doi:10.3102/00028312020002251

Blazar, D. (2015). Effective teaching in elementary mathematics: Identifying classroom practices that support student achievement. *Economics of Education Review*, *48*, 16-29. doi:10.1016/j.econedurev.2015.05.005

Blazar, D., & Pollard, C. (2017). Does test preparation mean low-quality instruction? *Educational Researcher*, *46*(8), 420-433. doi:10.3102/0013189X17732753

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008a, November). *Teacher preparation and student achievement* (NBER Working Paper No. 14314). Retrieved from http://www.nber.org/papers/w14314.pdf

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008b, May). *Who leaves? Teacher attrition and student achievement* (NBER Working Paper No. 14022). Retrieved from http://www.nber.org/papers/w14022.pdf

Boyd., D., Lankford, H., Loeb, S., & Wyckoff, J. (2010). Teacher layoffs: An empirical illustration of seniority versus measures of effectiveness. *Education Finance and Policy*, *6*(3), 439-454. doi:10.1162/EDFP_a_00041

Broatch, J., & Lohr, S. (2012). Multidimensional assessment of value added by teachers to real-world outcomes. *Journal of Educational and Behavioral Studies*, *37*(2), 256-277.

Casserly, M. (2015, October 30). A cap on the amount of testing time is the wrong answer for schools. *The Washington Post*. Retrieved from http://tinyurl.com/zrca4tn

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage.

Clotfelter, C., Ladd, H., & Vigdor, J. (2007a). *How and why teacher credentials matter for student achievement?* (NBER Working Paper No. 12828). Retrieved from http://www.nber.org/papers/w12828.pdf

Clotfelter, C., Ladd, H., & Vigdor, J. (2007b). *Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects* (NBER Working Paper No. 13617). Retrieved from http://www.nber.org/papers/w13617.pdf

Cohen, J., & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher*, *45*(6), 378-388. doi:10.3102/0013189X16659442

Darling-Hammond, L. (2003). Keeping good teachers: Why it matters and what leaders can do. *Educational Leadership*, *60*(8), 6-13.

Darling-Hammond, L. (2015). Can value-added add value to teacher evaluation? *Educational Researcher*, *44*(2), 132-137. doi:10.3102/0013189X15575346

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, *93*(6), 8-15. doi:10.1177/003172171209300603

Directions Evidence and Policy Research Group, LLP. (2014). *Ontario Regulation 274—Final report*. Retrieved from http://www.edu.gov.on.ca/eng/policyfunding/memos/nov2014/REG274EN.pdf

Doherty, K., & Jacobs, S. (2013). *Connect the dots: Using evaluations of teacher effectiveness to inform policy and practice. State of the states 2013*. Washington DC: National Council on Teacher Quality.

Fagginger Auer, M., Hickendorff, M., Van Putten, C., Beguin, A., & Heiser, W. (2016). Multilevel latent class analysis for large-scale educational assessment data: Exploring the relation between the curriculum and students' mathematical strategies. *Applied Measurement in Education*, *29*(2), 144-159. doi:10.1080/08957347.2016.1138959

Fantilli, R., & McDougall, D. (2009). A study of novice teachers: Challenges and supports in the first years. *Teaching and Teacher Education*, *25*(6), 814-825. doi:10.1016/j.tate.2009.02.021

Feldon, D. (2007). Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational Psychologist*, *42*(3), 123-137. doi:10.1080/00461520701416173

Frenzel, A., Goetz, T., Ludke, O., Pekrun, R., & Sutton, R. (2009). Emotional transmission in the classroom: Exploring the relationship between teacher and student enjoyment. *Journal of Educational Psychology*, *101*(3), 705-716. doi:10.1037/a0014695

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *The Review of Economics and Statistics*, *89*(1), 134-150. doi:10.1162/rest.89.1.134

Grissom, J., & Loeb, S. (2017). Assessing principals' assessments: Subjective evaluations of teacher effectiveness in low- and high-stakes environments. *Education Finance and Policy*, *12*(3), 369-395. doi:10.1162/EDFP_a_00210

Hanushek, E., Kain, J., O'Brien, D., & Rivkin, S. (2005). *The market for teacher quality* (NBER Working Paper No. 11154). Retrieved from http://www.nber.org/papers/w11154.pdf

Harris, D., & Sass, T. (2009, September). *What makes for a good teacher and who can tell?* (CALDER Working Paper No. 30). Retrieved from http://www.urban.org/sites/default/files/alfresco/publication-pdfs/1001431-What-Makes-for-a-Good-Teacher-and-Who-Can-Tell-.PDF

Hattie, J. (2003, October). *Teachers make a difference: What is the research evidence?* Paper presented at the Australian Council for Educational Research conference, Melbourne, Australia. Retrieved from http://research.acer.edu.au/research_conference_2003/4/

Hiebert, J., Morris, A., Berk, D., & Jansen, A. (2007). Preparing teachers to learn from teaching. *Journal of Teacher Education*, *58*(1), 47-61. doi:10.1177/002248710629526

Hill, H., Blazar, D., & Lynch, K. (2015). Resources for teaching: Examining personal and institutional predictors of high-quality instruction. *AERA Open*, *1*(4), 1-23. doi:10.1177/2332858415617703

Huang, F., & Moon, T. (2009). Is experience the best teacher? A multilevel analysis of teacher characteristics and student achievement in low performing schools. *Educational Assessment Evaluation and Accountability*, *21*(3), 209-234. doi:10.1007/s11092-009-9074-2

Ingersoll, R., & Strong, M. (2011). The impact of induction and mentoring programs for beginning teachers: A critical review of the research. *Review of Educational Research*, *81*(2), 201-233. doi:10.3102/0034654311403323

Ingvarson, L., & Rowley, G. (2017). Quality assurance in teacher education and outcomes: A study of 17 countries. *Educational Researcher*, *46*(4), 177-193. doi:10.3102/0013189X17711900

Jacob, A. (2012). Examining the relationship between student achievement and observable teacher characteristics: Implications for school leaders. *International Journal of Educational Leadership Practices*, *7*(3), 1-13.

Jepsen, C., & Rivkin, S. (2009). Class size reduction and student achievement: The potential tradeoff between teacher quality and class size. *The Journal of Human Resources*, *44*(1), 223-250. doi:10.1353/jhr.2009.0008

Kane, T., McCaffrey, D., Miller, T., & Staiger, D. (2013). *Have we identified effective teachers? Validating measures of effective teacher using random assignment* (MET Project research paper). Retrieved from https://eric.ed.gov/?id=ED540959

Kane T., Rockoff, J., & Staiger, D. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*(6), 615-631. doi:10.3386/w12155

Kersting, N., Givvin, K., Thompson, B., Santagata, R., & Stigler, J. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, *49*(3), 568-589. doi:10.3102/0002831212437853

King Rice, J. (2010). *The impact of teacher experience: Examining the evidence and policy implications* (CALDER Policy Brief 11). Retrieved from http://www.caldercenter.org/publications/impact-teacher-experience-examining-evidence-and-policy-implications

König, J., & Kramer, C. (2016). Teacher professional knowledge and classroom management: On the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME). *ZDM Mathematics Education*, *48*, 139-151. doi:10.1007/s11858-015-0705-4

Kraft, M., & Gilmour, A. (2017). Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher*, *46*(5), 234-249. doi:10.3102/0013189X17718797

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM Mathematics Education*, *40*, 873-892. doi:10.1007/s11858-008-0141-9

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, *105*(3), 805-820. doi:10.1037/a0032583

Kunter, M., Tsai, Y., Klusmann, U., Brunner, M., Krauss, S., & Baumert, J. (2008). Students' and mathematics teachers' perceptions of teacher enthusiasm and instruction. *Learning and Instruction*, *18*(5), 468-482. doi:10.1016/j.learninstruc.2008.06.008

Lampert, M., Franke, M., Kazemi, E., Ghousseini, H., Turrou, A., Beasley, H., Cunard, A., & Crowe, K. (2013). Keeping it complex: Using rehearsals to support novice teacher learning of ambitious teaching. *Journal of Teacher Education*, *64*(3), 226-243. doi:10.1177/0022487112473837

Levy, S.-A. (2015, August 16). Rule stymies teacher hiring in Ontario. *Toronto Sun*. Retrieved from http://torontosun.com/2015/08/16/rule-stymies-teacher-hiring-in-ontario/wcm/fdcbf26a-6fb6-4036-a806-70b490d13817

Mihaly, K., McCaffrey, D., Staiger, D., & Lockwood, J. (2013). *A composite estimator of effective teaching.* Arlington VA: RAND Corporation.

Morgan, C. (2017). Urge legislators to support teacher retention effort. *Naples Daily News*, p. 9B.

Mulholland, L., & Berliner, D. (1992, April). *Teacher experience and the estimation of student achievement*. Paper presented at the annual meeting of the AERA, San Francisco, CA. Retrieved from http://files.eric.ed.gov/fulltext/ED348350.pdf

Noerager Stern, P., & Porr, C. (2011). *Essentials of accessible grounded theory.* Walnut Creek CA: Left Coast Press.

Nye, B., Konstantopoulos, S., & Hedges, L. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237-257. doi:10.3102/01623737026003237

Ontario Ministry of Education. (2006). *New Teacher Induction Program: Induction elements manual.* Toronto, ON: Queen's Printer for Ontario.

Ontario Ministry of Education. (2012, September). Fair and transparent hiring practices. *Principals Want to Know*, 22. Retrieved from http://www.edu.gov.on.ca/eng/policyfunding/leadership/pdfs/issue22.pdf

Ontario spending millions to make education more inclusive. (2017, September 7). *Toronto Sun*. Retrieved from www.torontosun.com/2017/09/07/ontario-spending-millions-to-make-education-more-inclusive

Palardy, G., & Rumberger, R. (2008). Teacher effectiveness in First Grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis*, *30*(2), 111-140. doi:10.3102/0162373708317680

Palmer, D., Stough, L., Berdenske, T., & Gonzales, M. (2001, April). *Identifying teacher expertise: An examination of researchers' decision making*. Paper presented at the annual meeting of the AERA, Seattle, WA.

Richards, C., & Farrokhnia, F. (2016). Optimizing grounded theory for policy research: A knowledge-building approach to analyzing WTO e-commerce policies. *International Journal of Qualitative Methods*, *15*(1), 1-23. doi:10.1177/1609406915621380

Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrics*, *73*(2), 417-458. doi:10.1111/j.1468-0262.2005.00584.x

Rockoff, J. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *NEA Papers and Proceedings*, *94*(2), 247-252. doi:10.1257/0002828041302244

Ruzek, E., Domina, T., Conley, A., Duncan, G., Karabenick, S. (2015). Using value-added models to measure teacher effects on students' motivation and achievement. *Journal of Early Adolescence*, *35*(5-6), 852-882. doi:10.1177/0272431614525260

Santagata, R., & Guarino, J. (2011). Using video to teach future teachers to learn from teaching. *ZDM Mathematics Education*, *43*(1), 133-145. doi:10.1007/s11858-010-0292-3

Schön, D. (1983). *The reflective practitioner: How professionals think in action*. New York, NY: Basic Books.

Seebruck, R. (2015). Teacher quality and student achievement: A multilevel analysis of teacher credentialization and student test scores in California high schools. *McGill Sociological Review*, *5*, 1-18.

Staiger, D., & Rockoff, J. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives*, *24*(3), 97-118. doi:10.1257/jep.24.3.97

Stronge, J. (2010). *Effective teachers = Student achievement: What the research says.* Larchmont, NY: Eye On Education.

Sutcliffe, A. (2016). Grounded theory: A method for practitioner research by educational psychologists. *Educational and Child Psychology*, *33*(3), 44-54.

True, C., Butler, K., & Sefton, R. (2011). Substitute teachers: Making lost days count. *The International Journal of Educational Leadership Preparation*, *6*(1), 1-10.

Weisberg, D., Sexton, S., Mulhern, J. & Keeling, D. (2009). *The Widget Effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: The New Teacher Project.

Weiss, E. (1999). Perceived workplace conditions and first-year teachers' morale, career choice commitment, and planned retention: A secondary analysis. *Teaching and Teacher Education*, *15*, 861-879. doi:10.1016/S0742-051X(99)00040-2

Wente, M. (2013, June 6). Overeducated and underemployed: The teachers college mess. *The Globe and Mail*. Retrieved from http://www.theglobeandmail.com/opinion/wente-overeducated-and-underemployed-the-teachers-college-mess/article12366486/

Wilson, M., Hallam, P., Pecheone, R., & Moss, P. (2014). Evaluating the validity of portfolio assessments for licensure decisions. *Education Policy Analysis Archives*, *22*(6), 1-30. doi:10.14507/epaa.v22n6.2014

Winters, M. (2011). *Measuring effectiveness: Credentials unrelated to student achievement* (Manhattan Institute for Policy Research Issue Brief No. 10). Retrieved from http://www.manhattan-institute.org/pdf/ib_10.pdf