



“Who are these for? Is this for the teacher?”: Understandings of Expertise and Evaluation in the Era of ESSA

Madeline W. Good

University of Missouri - Columbia

United States

Citation: Good, M. W. (2023). “Who are these for? Is this for the teacher?”: Understandings of expertise and evaluation in the era of ESSA. *Education Policy Analysis Archives*, 31(55).

<https://doi.org/10.14507/epaa.31.7807>

Abstract: Understandings of teacher expertise in the US have transformed over the past 40 years, arguably being “narrowed” and “numericized” due to high-stakes accountability and neoliberal education reform movements. While this trend has been thoroughly studied under the No Child Left Behind Act (NCLB) and Race to the Top regimes, less consideration has been given since the Every Student Succeeds Act (ESSA) replaced NCLB in 2015, which notably pivoted away from some of the most stringent accountability practices of the previous era. This paper begins a new line of inquiry into teacher expertise in our current federal policy context, especially considering how understandings of expertise are constructed in school districts that never adopted the most high-stakes evaluation measures. By relying on Jessica Holloway’s (2021) technologies of risk management, this paper explores how teachers in one school understand expertise, focusing specifically on how evaluation and assessment technologies engage with and influence these understandings. Ultimately, it was found that teachers in fact held a plurality of understandings, yet complex and sometimes conflicting influences of assessment and evaluation practices also emerged. This paper argues that although risk management technologies have become commonplace, scholars and practitioners alike should continue to scrutinize their use under ESSA, particularly considering how they are being used, who they primarily benefit, and what consequences come from our reliance upon them.

Keywords: teacher expertise; teacher evaluation; accountability; supervision; Every Student Succeeds Act (ESSA)

“¿Para quién son estos? ¿Es esto para el maestro?”: Comprensión de la experiencia y la evaluación en la era de ESSA

Resumen: La comprensión de la experiencia docente en los EE.UU. se ha transformado en los últimos 40 años, posiblemente siendo “restringida” y “numerizada” debido a la rendición de cuentas de alto riesgo y los movimientos de reforma educativa neoliberal. Si bien esta tendencia se ha estudiado minuciosamente en virtud de los regímenes de la No Child Left Behind Act (NCLB) y Race to the Top, se le ha dado menos consideración desde que la Every Student Succeeds Act (ESSA) reemplazó a la Ley NCLB en 2015, que se alejó notablemente de algunos de los regímenes más estrictos. prácticas de rendición de cuentas de la época anterior. Este documento inicia una nueva línea de investigación sobre la experiencia docente en nuestro contexto de política federal actual, especialmente considerando cómo se construye la comprensión de la experiencia en los distritos escolares que nunca adoptaron las medidas de evaluación más importantes. Al basarse en las tecnologías de gestión de riesgos de Jessica Holloway (2021), este documento explora cómo los maestros de una escuela entienden la experiencia, centrándose específicamente en cómo las tecnologías de evaluación y evaluación emergen e influyen en estos entendimientos. En última instancia, se encontró que los docentes de hecho tenían una pluralidad de comprensiones, pero también surgieron influencias complejas y, a veces, conflictivas de evaluación y prácticas de evaluación. Este documento argumenta que, aunque las tecnologías de gestión de riesgos se han vuelto comunes, tanto los académicos como los profesionales deben continuar examinando su uso bajo ESSA, particularmente considerando cómo se usan, a quiénes benefician principalmente y qué consecuencias se derivan de nuestra confianza en ellos.

Palabras-clave: experiencia docente; evaluación docente; accountability; supervisión; Every Student Succeeds Act (ESSA)

“Para quem são estes? Isso é para o professor?”: Entendimentos sobre perícia e avaliação na era da ESSA

Resumo: A compreensão da experiência do professor nos EUA se transformou ao longo dos últimos 40 anos, possivelmente sendo “estreitada” e “numerizada” devido à responsabilidade de alto risco e aos movimentos de reforma educacional neoliberal. Embora essa tendência tenha sido exaustivamente estudada sob os regimes No Child Left Behind Act (NCLB) e Race to the Top, menos consideração foi dada desde que o Every Study Succeeds Act substituiu o NCLB em 2015, que se afastou notavelmente de alguns dos mais rigorosos práticas de responsabilidade da era anterior. Este artigo inicia uma nova linha de investigação sobre a especialização do professor em nosso atual contexto de política federal, especialmente considerando como os entendimentos de especialização são construídos em distritos escolares que nunca adotaram as medidas de avaliação de alto risco. Baseando-se nas tecnologias de gerenciamento de risco de Jessica Holloway (2021), este artigo explora como os professores de uma escola entendem a especialização, concentrando-se especificamente em como a avaliação e as tecnologias de avaliação emergem e influenciam esses entendimentos. Em última análise, constatou-se que os professores, de fato, tinham uma pluralidade de entendimentos, mas também emergiram influências complexas e às vezes conflitantes da avaliação e das práticas avaliativas. Este artigo argumenta que, embora as tecnologias de gerenciamento de risco tenham se tornado comuns, acadêmicos e profissionais devem continuar a examinar seu uso

sob a ESSA, especialmente considerando como estão sendo usadas, quem elas beneficiam principalmente e quais consequências advêm de nossa confiança nelas.

Palavras-chave: expertise docente; avaliação de professores; accountability; supervisão; Every Student Succeeds Act (ESSA)

“Who are these for? Is this for the teacher?”: Understandings of Expertise and Evaluation in the Era of ESSA

For nearly 40 years the U.S. federal government has played an increasingly influential role in public education. From the 1983 publication of *A Nation at Risk* through the 2009 federal grant competition Race to the Top (RttT), U.S. schools have gone through a transformational process resulting from standardization and accountability movements. While this has arguably impacted all aspects of public education today, there have been particularly influential changes in how we come to understand teacher expertise. Critical scholars since the early 2000s (Ball, 2003; Clarke, 2013; Holloway, 2021; Holloway & Brass, 2018; Jeffrey, 2002; Wilkins, 2011) have argued that in contexts where high-stakes and neoliberal educational accountability movements have attained dominance, teachers have been increasingly governed by “performative technologies” (Englund & Gerdin, 2019, p. 502) that are implemented with the goal of making teachers “visible, calculable, and comparable” (Clarke, 2013, p. 230). Through neoliberal logics, these changes in teacher accountability policies have been framed as strategies implemented to ensure access to quality teachers, increase student assessment scores, and help solve the U.S. crisis in education.

Resisting these often uncritically accepted logics, Holloway (2021) argues that in the most high-stakes environments, such policies frame teachers as risks that need to be managed, as opposed to professionals to be trusted, by utilizing an “onto-epistemic regime” that only allows for narrowed and rigid understandings of expert teachers and teaching to be legitimized (p. 99). This has led to an ongoing erosion of trust and personal agency in the de-professionalized or re-professionalized field of teaching (Holloway & Brass, 2018; Milner IV, 2013). Yet the consequences of such policies do not only impact teachers, as there are further concerns when we consider how heavy reliance on standardized assessments consequently narrow and numericize how we understand students (Bradbury & Roberts-Holmes, 2018; Daliri-Ngametua, 2022). This especially impacts students with historically marginalized identities as standardized assessments are deeply rooted in white supremacy and colonialism (Au, 2009; Viruru, 2009).

Importantly, however, we find ourselves today in a notably different and more flexible federal policy context under the Every Student Succeeds Act (ESSA), which has created avenues for more state control regarding how teachers are evaluated (Darling-Hammond et al., 2016). Considering new possibilities in teacher evaluation is essential, as preliminary evidence has found that the federal push for high-stakes teacher evaluation has been largely unsuccessful when considering student achievement, high school graduation rates, and college enrollment (Bleiberg et al., 2021). Yet Levinson, et al. (2009) remind us that “policy typically serves to reproduce existing structures of domination” (p. 769), challenging us to consider if new ways of constructing teacher expertise are being conceptualized under ESSA or if the influence of high-stakes evaluation under NCLB and RttT lingers on (Holloway, 2021).

The purpose of this exploratory qualitative study is to investigate how teachers in one school under ESSA understand teacher expertise using the following research questions: (1) How do teachers within a rural primary school community understand teacher expertise? (2) In what ways do evaluation and student assessment policies inform their understandings? The focus of this study will be one rural primary school in Missouri due to its particular policy context. Missouri has never

adopted the most high-stakes teacher evaluation methods and gives districts the opportunity to design their own teacher evaluation protocol, which the district included in this study has taken advantage of. The school also does not participate in the state-designed assessments that begin in third grade as it only serves preschool through second grade and its rurality may help the district “resist the pull of standardization” more than schools in more urbanized contexts (Tieken, 2014, p. 24). Overall, this school provides a key study site due to its avoidance of the most high-stakes evaluation practices, giving us an opportunity to explore whether the rigid ways of “knowing and being a teacher” have potentially ended (Holloway, 2021, p. 100) or if its ideological influence maintains a stronghold within the current era of teacher evaluation.

Federal Education Policy and Teacher Evaluation

President Ronald Reagan famously aimed to dismantle the U.S. Department of Education during his tenure, believing that the federal government had become too influential in U.S. school governance (Hayes, 2004; Mehta, 2015). His initial education policy agenda focused on local control, school prayer, and tax credits, yet after *A Nation at Risk* was published in 1983 under the supervision of the U.S. Department of Education, a new public and political debate emerged. Concerns about the “rising tide of mediocrity” in U.S. schools suddenly became top priority, and education reform jumped to the top of political agendas across party lines (Hayes, 2004; NCEE, 1983, p. 5). Some of the proposed solutions focused on content and standards, such as reprioritizing subjects like math and science so the U.S. would be more competitive on a global stage. Others, however, brought the spotlight onto teachers. The report argued that teacher salaries must become more competitive and qualifications for those wanting to enter the profession should increase in rigor to ensure that teachers are no longer “drawn from the bottom quarter of graduating high school and college students” (Mehta, 2015; NCEE, 1983, p. 123). Although recommendations within the publication were not universally adopted and implemented, *A Nation at Risk* constructed a new crisis in education, with teachers framed as being key players in the race to save U.S. schools.

A wave of local and state reforms emerged after 1983, including the standards-based movement under the Clinton administration, but the next highly influential education policy enacted specifically from the federal level was the 2001 No Child Left Behind Act ([NCLB]; Guthrie & Springer, 2004). A widely popular bipartisan piece of legislation at the time of its adoption, NCLB created the school accountability discourse still present across the U.S. educational landscape today. Instead of being guaranteed federal funding based on the previous compensatory approach, states and districts now had to comply with NCLB’s mandates and achievement expectations to access billions of dollars in funding. These included stipulations around the “highly qualified teacher”, as seen in the act’s Title II, which tightened expectations around teacher certification, subject-matter knowledge, and best practices in teaching (Cochran-Smith & Lytle, 2006). Some proposed that these new guidelines had the potential to increase the overall quality of the teacher workforce (e.g., Birman et al., 2007). Others, however, problematized the ways in which they narrowed what was deemed important or best regarding subject matter and instructional practices, arguing that NCLB framed teachers as the linchpin that will improve student achievement and save failing public schools (Cochran-Smith & Lytle, 2006). No matter how the potential outcomes were perceived, a new framework for identifying expert, or “highly qualified,” teachers was inevitably created under NCLB.

At the peak of the NCLB-era came Race to the Top (RttT) in 2009 under President Barack Obama, which was a \$4.35 billion dollar federal grant competition designed to “incentivize excellence” in U.S. public schools (Viteritti, 2013, p. 2102). The RttT reform movement proved to be strikingly influential in the realm of teacher quality, as it was the first time teacher evaluation

scores were required to be partially based on student assessment results for districts to gain federal funding (Baker et al., 2013). This led to widespread use of “calculation, inspection, intervention, and audit” to improve teacher quality, with heavy reliance on controversial value-added models (VAMs) which use advanced statistical techniques to measure how much a teacher influences their students’ learning (Holloway, 2021, p. 31). Ultimately, 47 states plus Washington D.C. applied for the grants (Finch, 2017). Although fewer than half of the applicants received funding, the influence of the RttT requirements was seen across the US, as even states that never applied for grant support made changes to stay aligned with the educational reform trends (Howell, 2015).

Near the end of NCLB and RttT, it became increasingly clear that the legislations’ lofty goals were unattainable for most schools. After granting states waivers in 2011 for those who were unable to meet the policies’ expectations, the Every Student Succeeds Act (ESSA) was eventually passed, replacing NCLB in 2015 (Darling-Hammond et al., 2016; Viteritti, 2013). ESSA changed much of the high-stakes tenor that NCLB had implemented, providing more state-level flexibility in designing school oversight, opening opportunities for a “more holistic approach” to teacher evaluation (Darling-Hammond et al., 2016, p. 1). Notable shifts occurred in how state teacher evaluation systems were designed, with many moving away from some of the most high-stakes accountability practices, such as value-added modeling (Close et al., 2020). While there are still states, schools, and districts relying on high-stakes teacher evaluation systems today, across the larger U.S. landscape, that trend has been declining and there appears to be some new flexibility regarding how states and districts can define teacher expertise.

What Does it Mean to be an Expert Teacher?

When “teacher expertise” has been studied in educational research and developed in educational policy over the past 20 years, the concepts of teacher quality, teacher effectiveness, and/or teacher expertise have most frequently been used to consider how high-quality, effective or expert teachers can be differentiated from their colleagues (e.g., Berliner, 2001; Darling-Hammond, 2000; Palmer et al., 2005). In general, three measurements have been used to make such value judgments: inputs, processes, and/or outputs (Goe et al., 2008). Inputs are characteristics that teachers bring to the profession, like educational training, personal experiences, and certifications. Processes are what occurs within the classroom, including how a teacher interacts with their content, pedagogy, and students. Outputs are measurements of the teacher’s impact, such as student test scores and survey data. Input and process characteristics have historically been used to identify “high-quality” or “effective” teachers (e.g., Darling-Hammond, 2000), yet after NCLB and RttT, policies across the US increasingly identified effective teachers using output characteristics that focused heavily on student standardized assessment scores (Goe et al., 2008). The implementation of “performative technologies” such as audits, performance measures, and rankings to assess such output characteristics were key for many educational reformers (Englund & Gerdin, 2019, p. 502). They provided an avenue for teachers to be objectively held accountable for what was increasingly deemed the most important part of their jobs: showing clear evidence of student gains on standardized assessments (Ball, 2003; Clarke, 2013).

While these narrowed and datafied understandings of teacher expertise gained prominence, critical scholars have voiced concerns about their implications on teachers and schools (see e.g., Ball, 2003; Clarke, 2013; Holloway, 2021; Holloway & Brass, 2018; Jeffrey, 2002; Wilkins, 2011). When considering the US specifically, Holloway (2021) found that in one of the most high-stakes evaluation contexts under NCLB and RttT, expert teachers were increasingly differentiated from their non-expert colleagues through the reliance on metrics such as student academic performance

and rigid definitions of instructional best practices. Holloway (2021) argues that these types of evaluation policies and processes construct an “onto-epistemic regime” that “constitutes the teacher *being* through targeted training, data instruments, and data outputs” while also providing the “*knowledge* of teacher quality, professionalism, and value” (p. 104). In these contexts, the construction of the perpetually comparable and auditable teacher was typically accepted, with no overwhelming tension between how high-stakes performance policies (VAMs, observation rubrics, merit-pay, etc.) identified “good” teaching and what teachers themselves personally understood to be “good” teaching (Holloway, 2021; Holloway & Brass, 2018). In fact, critiquing any part of the process put you at risk for being viewed as an unacceptable teacher who was unfit for the school.

In the more recent U.S. federal policy context under ESSA, it appears that there may be space for more nuanced understandings of teacher expertise to be constructed in policy and practice, but little work has been done to explore if and how this “onto-epistemic regime” has morphed or evolved in a post-NCLB era. Additionally, much of the work around teacher evaluation in the US since NCLB was implemented has focused on teachers being evaluated with the most controversial and/or high-stakes models (e.g., Amrein-Beardsley, 2008; Ford et al., 2017; Hallinger et al., 2014; Lavigne, 2014), leaving states that never adopted such practices potentially under-investigated. This study aims to begin a new line of inquiry to understand how teacher expertise is being constructed and understood in a context that has likely been overlooked as high-stakes teacher evaluations took hold across the US.

Theoretical Framework

This study’s theoretical framework utilizes the work of Holloway (2021) as it focuses on the ways in which high stakes teacher evaluation policies help create ‘onto-epistemic regimes of truth’ that construct and legitimize rigid, standardized, and datafied conceptualizations of what it means to be a teacher. According to this “regime of truth”, “acceptable” teachers are those who embrace high-stakes evaluation and surveillance, viewing themselves as perpetually “imperfect” and “auditable” while always striving “to become better versions of themselves” (Holloway, 2021, p. 100). While evaluation, metrics, and standardization are not “inherently bad,” Holloway (2021) argues that “the problem emerges when accountability tools and techniques supersede teacher expertise, training and experience, and thus undermine the value of professional discretion that is necessary for responding to the immediate needs of students and schools” (p. 159). To understand how these regimes of truth are created, Holloway conceptualized five technologies of risk management (see Table 1) used to mitigate the potential “risk” that teachers brought into the classroom (2021, p. 49). These five technologies include numericization, surveillance, normalizing judgments, examination, and discipline.

The first technology, numericization, refers to how teachers are “made knowable as objects of knowledge” through numbers (Holloway, 2021, p. 50). This is based on the work of Rose (1999) who framed numbers as a governance technology that has “achieved an unmistakable political power” due to their ability to determine who holds power in a political system, be used as diagnostic instruments, make modes of government judgeable, and allow complex state institutions to govern in a modern world (pp. 197-198). In Holloway’s (2021) framework, numbers are primarily used in value-added models and observation rubrics to make teachers “knowable” numerically. The next three technologies of risk management include surveillance, normalizing judgments, and examination (Holloway, 2021). These three technologies are based on the work of Foucault (1977) in his widely read and studied text, *Discipline and Punish*. In this work, Foucault analyzes how “individuals are ‘administered’ by the various bureaucratic institutions [...] that increasingly render

selves docile in the modern world” within an 18th-century carceral system (Leitch et al., 2018, p. 1388). Foucault (1977) unveils an image of governmental practices that “render selves docile” through constant surveillance, or hierarchical observation, normalizing judgments “making it possible to measure gaps [and] to determine levels,” and the combination of the two which leads to examinations (p. 184). In Holloway’s (2021) framework, surveillance focuses on how teachers are made both explicitly and implicitly visible through observations, lesson plan submission, and evaluation conferences. Normalizing judgments occurs through the legitimization of evaluation policies and observation rubrics, making teacher-to-teacher comparison possible and necessary. Then by combining surveillance and normalizing judgments, examination is imparted via the implementation of the evaluation process.

Table 1

Technologies of Risk Management

Technology	Function	Practice/Instrument
Numericization	The process of turning matters into numbers - making teachers knowable as objects of knowledge (Rose, 1999)	Rubrics Value-added models
Surveillance	The making of teachers, as well as teachers’ practices and attributes, visible, both explicitly (e.g., observations) and implicitly (e.g., making their thinking visible) (Foucault, 1977)	Observations Lesson plan submission Pre- and post-conferences Data dashboards
Normalizing Judgements	The setting of a standard or normal way of making judgments about teachers so that comparisons can be made about them (Foucault, 1977)	Rubrics Evaluator forms/guidelines Evaluator training handbook Evaluator conference prompts Evaluator training course
Examination	The combination of surveillance and normalized judgements - teachers are made visible and then examined based on normalized judgements (Foucault, 1977)	Rubrics Value-added models Observations Lesson plan submission Pre- and post-conferences Data dashboards
Discipline	The combination of numericization, surveillance, normalizing judgments and examination that disciplines teachers to conduct themselves in desired ways (Foucault, 1977)	Evaluation outcome use Cluster meetings/training Self-reflections Pre- and post-conferences Observations Rubrics Value-added models

Note. From *Metrics, standards, and alignment in teacher policy: Critiquing fundamentalism and imagining pluralism*, by J. Holloway, 2021, p. 50. Springer Singapore. Reprinted with permission.

Through the combination of the first four technologies of risk management, teachers are then “disciplined” to conduct themselves in desired ways (Holloway, 2021, p. 50). In this process, teachers are rendered docile through “individualization” while power “becomes more anonymous and more functional” (Foucault, 1977, p. 193; Holloway, 2021). This is where “acceptable” ideas of a teacher are formed, feeding into an “onto-epistemic regime” that dictates what behaviors, attitudes, and actions must be required in teacher evaluations to protect us all from the threat of the risky teacher (Holloway, 2021).

Research Design & Methodology

This project is an exploratory qualitative study (Stebbins, 2001) designed to research how teachers in one school community understand teacher expertise, representing the beginning of a long-term investigation into the development of teacher expertise during the era of ESSA. By prioritizing flexibility, as opposed to being limited by issues of “sampling” and “validity” that will eventually be addressed over the course of multiple studies, I am able to approach the idea of teacher expertise with exploratory openness, which is essential for “ideas that have just been brought to light” (Stebbins, 2001, p. 5).

Study Site and Teacher Evaluation Context

This study’s site is a rural, predominantly white primary school, which I refer to as Forest Elementary¹. Located in Missouri, it serves grades pre-kindergarten through second grade in the Cartersville Public School District ([CPSD]; see Table 2). While CPSD is technically situated within a “remote town” based on NCES locale classifications (NCES, 2020), it is one of the larger districts in the state based on square mileage and primarily serves students located in rural communities and on remote roads. Additionally, Forest Elementary is considered a large school in the community as they serve over 500 students from preschool through second grade.

Table 2

Demographic Information for the 2020-2021 School Year

	Cartersville Public School District	Forest Elementary
Enrollment (total)	3,873	522
White (%)	86.5	89.1
Black (%)	1.4	1.7
American Indian/Alaska Native (%)	0.3	*
Asian (%)	0.6	*
Hawaiian/Pacific Islander (%)	0.1	*
Hispanic (%)	7.5	5.4

¹ Pseudonyms are used in the study for school, district, and teacher names as well as the teacher evaluation and adaptive learning technologies described.

	Carterville Public School District	Forest Elementary
Multi-race (%)	3.6	3.3
Free and Reduced Lunch Population (%)	41.9	49.8

Note. Data from the Missouri Department of Elementary and Secondary Education. Asterisks (*) signify no data for that demographic category.

Regarding teacher evaluation, Missouri allows for each district to design and adopt their own desired model as long as it follows state guidelines. CPSD has taken advantage of this flexibility by creating their own system that aims to “enhance the learning process, establish a culture of learning and promote an environment of continuous growth” within its schools (CPSD, 2017, p. 4). There are four main phases of the evaluation system. First, teachers develop their growth plans at the beginning of the school year, choosing one of four indicators from the evaluation rubric to focus on: intentional planning, teaching, learning, and professionalism. They then choose a specific goal to reach within their chosen domain and articulate an action plan they will follow to reach that goal. Part two of the plan includes informal walkthroughs conducted by the principal as well as one to three (or more if needed) formative observations depending on how many years of experience you have in the district. For part three, each teacher meets with the building principal for a mid-point check where they review the progress made toward the identified goal, discuss relevant student data collected thus far during the school year, and touch base about other topics as needed. Lastly, the process ends with a summative evaluation at the end of the year where both the identified goal and student data are discussed again before the teacher is given their final evaluation performance review.

Participants and Data Collection

Participants were first recruited from Forest Elementary at a general faculty meeting where I shared my project with the staff and provided contact information for those interested in participating. Once I had my first participant, I then used snowball sampling to identify colleagues that might also be interested in the project (Savin-Baden & Major, 2013). One participant recommended that I contact a teacher who recently left Forest Elementary and got a job in a neighboring school district. Eventually, the same people were repeatedly recommended during snowball sampling, so participant saturation was assumed. Ultimately, seven current and one previous teacher at Forest Elementary participated in the study (see Table 3)². All teachers self-identified as white women, with one describing herself as Hispanic and white. They collectively had 14 years of teaching experience on average, ranging from one first year teacher to one teacher in her 23rd year. At the time of this study, they were each teaching kindergarten, first, or second grade.

² The participant who no longer worked at Forest Elementary was included in this study as they could potentially be more candid about their opinions of Forest Elementary and its principal Laura. Additionally, their recent experience outside the district could also illuminate the influence that local leaders have over evaluation policies and practices.

Table 3*Participant Information*

Participant	Position	Years of Teaching Experience	Sex	Race	Interview Fall 2021	Interview Fall 2022
Becky	2 nd Grade	11-15 years	Female	White	X	
Beth	2 nd Grade	21+ years	Female	White	X	
Sophie	2 nd Grade	6-10 years	Female	White	X	
Hillary	1 st Grade	1-5 years	Female	White	X	
Lisa	Kindergarten	16-20 years	Female	White	X	X
Sandra	Kindergarten	21+ years	Female	White	X	X
Patty	Kindergarten	21+ years	Female	White	X	X
Lauren	Kindergarten	1-5 years	Female	Hispanic/ White	X	
Laura	Principal	20; 10 years as principal	Female	White	X	X

Note. Teacher years of experience are shared in ranges to ensure anonymity.

Data for this study was primarily collected through two rounds of interviews with the classroom teachers. The first round of interviews was conducted in the fall of 2021 (see Table 3). Seven of these interviews were recorded and transcribed. One of the teacher interviews was not audio recorded due to in-the-moment technical difficulties, so instead their responses were recorded through typed notetaking over a phone call.³ Interviews lasted between 35 and 60 minutes.⁴ The interviews were semi-structured, relying on my research questions and theoretical framework to guide discussion topics (Brinkmann, 2014). These topics focused upon understandings of teacher expertise (e.g., “How do you understand teacher expertise? How do you identify expert teachers?”), how evaluation and student assessment influenced these understandings (e.g., “How much do your evaluations influence how you understand what “expert” teaching is?”), and if there were any points of dissonance between how they conceptualized expertise as compared to their colleagues, educational leaders, and community members (e.g., “Do you feel any tension between how you understand teacher expertise and how your principal understands it?”). As I learned more about the context over the course of the interviews, I adapted and refined interview questions appropriately, especially when topics pertained to aspects of the study’s theoretical framework. This occurred when I learned about teachers needing to submit lesson plans weekly (e.g., surveillance) and when it

³ Data collected from notetaking is notably different than the rich data garnered from interview transcripts. It was apparent that this was an important perspective to include since they were the only novice teacher recommended during snowball sampling. Thus, while acknowledging the limitations, this interview was still included in the study.

⁴ Two of the teachers were only able to be interviewed during their planning periods due to school and district leadership positions as well as caretaking duties before and after school, which limited the length of their interviews. While they still provided key insight for this study, this is a limitation of the data collected.

became apparent how influential BrightStar was to their teaching and evaluation (e.g., numericization).

During the fall of 2022, the second round of interviews were conducted. These follow-up interviews occurred after the principal reached out to let me know that she had left Forest Elementary and was now working in a different school setting. There had also been significant changes in local school board and district leadership since the interviews in the fall of 2021, so I decided to facilitate an additional round of interviews as new insights may emerge based on these developments. Ultimately, three teachers from the initial round of data collection were available and willing to be interviewed again, with these conversations lasting on average 45 minutes. The second round of interviews were more structured and focused primarily on four main topics that had emerged as key findings from the first round of interviews: how they understood teacher expertise, how these understandings aligned (or didn't) with the new administration, if/how the teacher evaluation process had changed, and if/how the heavy reliance on BrightStar had changed.

Due to the Covid-19 pandemic making classroom observations unfeasible and unsafe in the fall of 2021, information gathered from the interviews was additionally confirmed and triangulated with two interviews (fall of 2021 and 2022) conducted with the principal of Forest Elementary as well as artifacts related to the district teacher evaluation process. The principal interviews focused primarily on her views of teacher expertise, influence over the formal evaluation process, and relationship with the BrightStar program. Artifacts were also collected in the fall of 2021 and 2022, which included the rubric used to formally assess teachers, the rubric that connected student BrightStar scores to the teacher's evaluations, a working document used when district leaders were designing the evaluation in 2018/19, the district presentation about the evaluation process, and a publicly available hour-long video overview of the district evaluation background and structure.

Data Analysis

For data analysis, open and theoretical coding were used. For initial coding, I used a combination of descriptive and in vivo coding depending on which type of code captured participant voice and meaning most effectively (Saldaña, 2013; see Table 4). For the interview that was not recorded, I used descriptive coding to analyze the interview notes. The codes were then compiled into a separate document to identify emergent categories identified within each interview subtopic, such as "it depends", "generally non-influential", and "time consuming" (Savin-Baden & Major, 2013). For theory-based coding, the associated subtopics and their relevant initial codes were categorized (if possible) into the five "technologies of risk management" within Holloway's (2021) framework. Next, overarching themes, as well as contradictions, were then identified across the open and theoretical codes (Ellingson & Sotirin, 2020). Some themes, such as "Student Assessment Scores" were identified within most interviews and connected clearly to Holloway's (2021) technologies of risk management. Others, such as "it depends", were in most interviews as well, but did not necessarily fit within the theoretical framework, so they were only included in the open coding. Moments of dissonance were also noted, such as "questioning Brightstar", even if they only appeared in a few interviews, due to their particular relevance to Holloway's (2021) onto-epistemic regime. Overall, my data analysis was a highly iterative process, as I continually revisited the codes, interviews, and documents throughout analysis and writing.

Table 4*Open Codes for Data Analysis - Examples*

Interview Subtopic	Example Initial In Vivo Codes	Example Initial Descriptive Codes	Example Emergent Categories
Teacher expertise	“willing to change”	student engagement	“it depends”
Teacher evaluation	“authentic conversations”	doesn’t impact teaching	generally non-influential
Lesson plan submission	“who are these for?”	redundancy	time consuming
BrightStar	“drive my instruction”	student burnout	relied upon heavily

To ensure trustworthiness, I ground all findings in the data collected from the interviews (Lincoln, 2015). When a quote was identified as important based on my research questions or theoretical framework, I returned to the original recordings to verify my interpretation of its content before including it in the findings. I also ensured that both themes and their contradictions were included to avoid only discussing confirmatory findings. Additionally, as a white woman former teacher researching a predominantly white rural school context, my identity brings strengths and limitations to this work. While my positionality may help me relate to, understand, and build trust with participants, it also creates assumptions and blinders that I inevitably carry. My shared racial and gendered identities with most of the participants also risks further normalizing oppressive structures such as whiteness and gendered power dynamics in schools and educational research. While the purpose of this study is not to scrutinize the racialized and gendered nature of teacher expertise at Forest Elementary, recognizing their inevitable influence on both teaching and research hopefully disrupts this “normalization” and emphasizes the importance of continuing this work across a variety of contexts.

Findings

The findings from this exploratory qualitative study illustrate potential avenues for new constructions of teacher expertise in the era of ESSA as well as lingering influences of narrowed and numericized understandings of expertise. In this section I address the general findings behind the first research question, “How do teachers within a rural primary school community understand teacher expertise?”, which shows a plurality of understandings held at Forest Elementary. I then move to the second research question: “In what ways do evaluation and student assessment policies inform their understandings?” Here, three primary topics emerged, including perceptions of the “official” district evaluation tool, reflections on the school’s requirement of weekly lesson plan submission, and insights into the adaptive technology, Brightstar, used by all teachers. Ultimately, these findings explore the themes, contentions, and contradictions that emerged from the interviews, providing evidence of both the destabilization of Holloway’s “onto-epistemic regime” as well as its continued presence in Forest Elementary classrooms.

What Is Teacher Expertise?

Teachers understood expertise in a number of different ways, including collegiality, being in leadership positions, classroom management skills, relationship building with students, willingness to change, and student academic success, among other qualities (see Table 5). Ultimately, there was no singular skill or measure that all teachers relied upon to identify who the expert teachers were in their building. Some teachers emphasized how specific behaviors help differentiate the expert teachers from their non-expert colleagues. Kindergarten teacher Patty noted the importance of teachers sharing resources and information in her understanding of expertise, stating that, “it’s the ones that usually are sending out things that they’ve learned, or things that they’ve made, or things that they’ve created, and they’re sharing with the entire building.” Others, such as second grade teacher Sophie, instead emphasized how people’s strengths may influence who she sees as experts at different moments:

Teaching is an art and a science. So there are teachers like myself who really love the science part of it [...] But as far as that art piece, it’s kind of like knowing all the lyrics but not be able to carry a tune. So there’s different people I would go to depending on what I am looking for.

Table 5

Teacher Understandings of Expertise

Participant	Position	Example Codes
Becky	2 nd Grade	supports leadership; takes risks; self-confidence; “natural abilities”; respected by peers
Beth	2 nd Grade	“it changes”; “depending on the topic”; go to for help; “connections with families”; chosen for leadership positions
Sophie	2 nd Grade	“it depends”; student engagement; knowledge of child development; “grade level leaders”; “have an answer”;
Hillary	1 st Grade	humble; “willing to change”; adjusts to “classroom needs”; adaptable; uses data
Lisa	Kindergarten	builds relationships; student engagement; kind; “move that score”; “support other teachers”
Sandra	Kindergarten	“it depends”; subject knowledge; behavior management; “good rapport with students”; worked with coaches
Patty	Kindergarten	“see the results”; “general professionalism”; “commitment to teaching”; “willingness to show up”; “shares everything”
Lauren	Kindergarten	“willing to change”; relates to students; asks for help; technology knowledge; “not jumping on the newest thing”

Sophie’s reflection connects to the most common theme found when I asked about how participants understood teacher expertise: that definitions of teacher expertise were contingent upon what parts of teaching you were referring to. Beth, a second-grade teacher, stated that “it just depends on what the topic is.” She described how school administration may rely on more experienced teachers when working with issues surrounding curriculum, but then the expertise of

those same teachers may not be present when new programs are adopted and implemented. Sandra, a kindergarten teacher, also argued that it “depends on what we’re talking about,” stating that even different subjects have different experts because she looks to different people for support depending on the specific lesson she is planning. Overall, “it depends” can be best encapsulated in a quote from Lauren, the former Forest Elementary teacher who now teaches at a neighboring district:

There’s different areas of expertise [...] Are you an expert in relating to your kids?
Are you an expert with lesson instruction? Are you an expert with a program?
There’s all these different pieces of expertise that kind of feed into each other. And I think you can be an expert in different areas of teaching.

It is apparent that most of the teachers in this exploratory study did not see expertise as one specific metric or personal quality. Instead, it was understood as something contingent upon what you were specifically needing or looking for.

There were a few teachers who did not see expertise as such a flexible concept, framing it as more rigid and unchanging. For example, Becky, a second-grade teacher, specifically stated that she looks for “natural abilities. You got natural abilities. They’re probably leaders in other areas, like the academic area. [...] People that have those personality qualities tend to rise to the top in their professional areas.” Becky saw expertise as based on innate personality qualities or skills, contrasting with the view that “it depends” as articulated by her colleagues. Additionally, some teachers described expertise as a more flexible concept while simultaneously emphasizing the importance of student assessment scores as evidence of expertise. Kindergarten teacher Lisa saw qualities such as collegiality and student engagement as key characteristics of expert teachers, yet she also noted that “being able to move that [assessment] score does say a lot about a teacher.” Thus, while “it depends” shows a flexibility of thought regarding teacher expertise at Forest Elementary, it does not necessarily represent all the understandings shared in this exploratory study.

Interestingly, Laura, the Forest Elementary principal, held a more specified understanding of teacher expertise. While she shared that qualities such as “knowledge of the material” and “able to make adjustments” were important, it appeared that her primary understanding of teacher expertise was directly related to student learning and academic growth. She shared that teachers must, “see the importance of all of our students learning [...] that all of our students are making growth toward their own personal knowledge.” Laura framed student academic growth as an issue of “ethical responsibility” for teachers and schools, stating “if I’m in the classroom for a year with a student, I should grow them a year. Even if they start below or two years above, it’s my job to grow them.” She also noted that there are teachers she would consider “superior” whose students typically grow a year and a half by the end of the school year. Thus, in contrast with most of the teachers interviewed, it appears that Laura sees student academic growth as carrying more weight when understanding teacher expertise than other qualities, skills, or characteristics.

Sandra confirmed this perspective when reflecting upon differences she noticed with the school and district leadership changes that had occurred since her first interview:

I think our former administration [...] would have said our teacher experts are the one who have the data. And I think now, it would be the teacher experts are the ones who are... it’s more of a feeling. So your teacher experts are the ones where when you walk in the room, your students are engaged, and they’re respectful to each other, and they’re learning and they are progressing. They may not have the best scores, but they are progressing.

Sandra continued by sharing that the previous administration “were looking at data points, they were looking at graphs and numbers” more than considering what they were observing in the classroom. She then contrasts that with the new administration’s focus on qualities other than assessment data, stating, “some of the teachers that have been highlighted in conversations or in emails, they may not have the best test scores, and they would admit that, but they are, they are making gains with your students.” Ultimately, it appears that ideas about student “gains” and “growth” are now less directly connected to student assessment data, exemplifying how changes in local educational leadership may influence what is understood as teacher expertise at Forest Elementary.

Teacher Expertise, Evaluations, and Assessments

The study’s second research question further illuminated how teachers understood teacher expertise, especially pertaining to the topics of evaluation and student assessments. The three primary findings here focus on how the “official” district evaluation, required lesson plan submission, and student evaluation scores influence their conceptualizations of expertise in teaching.

Official Evaluation

Overall, teachers generally maintained a neutral to negative framing of the “official” district evaluation protocol. Beth described it as a “hoop to jump through” due to the delayed feedback you get from the evaluation program itself, sharing that “to be honest, by the time that we get the information from it [...] by then I’m finished” with the school year. Sandra echoed Beth’s sentiments, stating, “I’ve never felt like it is what impacts my teaching, positively or negatively [...] it gives me an area to focus on, but I don’t know if I necessarily get better at it.” Interestingly, Sandra, among other teachers, also mentioned how the scoring system in the evaluation document is poorly designed. She shares how it is virtually impossible to reach a 6 or a 7 (the two highest possible scores) because of their unattainable requirements:

Sometimes teachers, and I’m probably guilty of this, too, at the start of the year score ourselves lower than we need to because it’s almost impossible to reach that... Is it a six or a seven? And once you reach it, then what? Honestly, almost every year I feel like I’m at a five. Every single year.

Lisa also mentioned this issue, stating that they are specifically told that “no one will ever reach distinguished, so don’t even worry about that.” Thus, not only are the evaluation forms framed as a “hoop to jump through,” they are also designed in a way that makes it virtually impossible for teachers to achieve the highest score, leaving them in a situation where they are potentially trying to achieve a level of success that is unattainable.

While teachers such as Beth and Sandra see the evaluation tool as neutral at best, others such as Sophie see it as actively harmful. When asked about the influence of evaluation on teacher expertise, Sophie describes a more concerning perspective, especially for new teachers:

When you’re someone who... did great at high school, then did great at college, you just did great at everything. Then you get here and you work your darn hardest and can’t even, can’t even... find yourself out from under the water. [...] And so that just makes me sad for people like, I’m in a place where I’m like, all right I’ve got, I think I’ve got it figured out a little bit where I can at least breathe. I just feel bad for the brand-new ones that are like, “I’m terrible. I picked the wrong thing!”

Here, Sophie argues that the evaluation tool’s unreasonable expectations lead to feelings of demoralization, especially those new to the profession.

Lauren shared similar sentiments as Sophie, stating that although she loved her experience in Carterville, “there are things that fogged their good teachers because of the unrealistic expectations.” These expectations ultimately led her to search for employment in a neighboring district, which has turned out to be a much more positive experience for her:

I feel like the best teacher I’ve ever been simply because I can do what’s best for the kids in my classroom and I feel trusted to do that. I’m not constantly scared that someone is going to come in my room and evaluate me wrong because they’re [current administrators] constantly in and out looking. It’s just a night-and-day difference of stress level for sure.

Lauren shared that the unreasonable expectations in the Carterville evaluation protocol, along with the way it was specifically implemented at Forest Elementary, led her to struggle with constant anxiety for fear that she would be falsely evaluated in a negative way.

While most teachers framed the official evaluation tool as innocuous at best, there were a few teachers, such as Patty, who noted the positives that can come out of the evaluation process. She shared that her “principal tries her very best to make sure that it is beneficial [...] the evaluation system has allowed me time to have those moments with her [the principal]. But I wouldn’t say it’s directly from the evaluation.” While the evaluation policy itself is not framed in a positive light, the approach taken by the principal when engaging in the required conversations were. It is important to note that other teachers may not have made connections between the evaluation’s policies and meaningful conversations with their principal, even if they are linked. With the data collected, however, it appears that teachers either generally do not see the evaluation protocol as overly influential in their understandings of expertise or if they do, they see its connection to high expectations as unreasonable and potentially harmful.

Lesson Plan Submission

The most polarizing issue regarding the evaluation protocol at Forest Elementary pertained to the expectations surrounding weekly lesson plan writing and submission. Although it was not a policy enforced across all schools in CPSD, it was a district-level expectation that Forest Elementary followed. Based on these expectations, teachers should not be able to receive “proficient” in the “intentional planning” domain of their evaluation if they did not have detailed plans posted and available for review. At Forest Elementary specifically, teachers were expected to have the next week’s lesson plans completed, printed, and posted in their classrooms by Friday afternoon before they left for the weekend. Overall, this practice was controversial and there was no unified perspective about the benefits and drawbacks of its implementation.

There were many teachers who framed it as an overwhelming and unnecessary part of their evaluation process. Lisa took most issue with having to restate what was already articulated in their required curriculum:

I have a huge issue with having to put back down what our curriculum already says [...] Yes, you have to be intentionally planning, but goodness gracious... when I’m just copying exactly what the curriculum already says. That seems like a lot of time for something.

Sophie added to the concern around how much time lesson planning took, sharing that, I know people who worked here for years and have been in the same grade level for years spending 7 hours on a Sunday trying to get lesson plans done and they’re 25 pages long. And then when you get to that point, you’re like, ‘Who are these for? Is

this for the teacher?’ [...] There’s a lot of things that you have to have in your lesson plan and a lot of things that you have to have posted. A lot of things... a lot of things that mean nothing for the teacher and it’s more for someone looking in.

Sophie is overtly critical not only of the lesson plan requirements, but also with the idea that it provides benefit only to those who are “looking in.” She later shares that unless you are a new teacher, your first evaluation is entirely based on what your submitted lesson plans look like, stating, “that’s your worth as the teacher is that one snapshot in time. And I think that is a hard pill to swallow.”

Lauren was also clear about her critique of the lesson plan submission requirement, stating that it was one of the primary reasons she left CPSD. When describing the administration in her new position, she shares:

So they have never looked at my plans. Honestly. They have never asked me for my plans. They’ve never looked for my plans. But I feel more planned out here than I ever did at [Carterville] because I don’t have to attach all this, you know, it’s not so tedious. And so to me, my plans are for me to know where I’m going with my teaching.

The freedom in designing lesson plans for herself, as opposed to her administration or evaluation protocol, has led her to feel “more planned out” than she ever has before. She says she now feels “like ten times the teacher I was able to be [in Carterville]” because she is, somewhat paradoxically, “able to prepare more.” Overall, it is apparent that Lauren, along with many of her former colleagues, felt like the required lesson plan submission was at best time consuming and at worst anxiety-inducing.

Importantly, there were teachers that did not take issue with the lesson plan requirements. Hillary, a new teacher in first grade, found them reasonable because she said she would be completing lesson plans in a similar format anyway. She shared that she “likes that accountability” as compared to her more experienced peers and appreciates that she can “lean on them” if she ever needed to justify her instructional practices. She did note, however, that there is a chance her perspective will change once she has a few more years of experience, so it may become more of a hoop to jump through as framed by her peers. Patty also mentioned the potential benefits of this lesson plan submission requirement for new teachers specifically, sharing that “if I had been required to do this my first five years of teaching, it would have been much more helpful.” She stated that she “didn’t learn how to effectively plan until I came to this district,” and appeared to be grateful for the explicit emphasis that CPSD and Forest Elementary put on lesson planning. Beth also did not take issue with the lesson plan submission, stating that, “I feel like... I think probably because I’ve done it for so long, it’s just kind of automatic [...] it’s just kind of ingrained in me that I need to make sure I have these things.” While Beth did not specifically frame them as beneficial, they appeared to be less of a burden for her due to her experience in the profession.

Notably, Sandra, Patty, and Lisa shared that under the new principal, the expectations around the lesson plan submission had been adjusted. They still needed to have their plans available online for their principals to view, but they no longer had to print and post them by Friday at the end of the school day. Additionally, they did not have to be as detailed as before, with Lisa sharing that her weekly lesson plans went from being “about forty pages to about twelve.” All three teachers, including those who had not shared negative feelings about the lesson plan requirements in the first round of interviews, responded positively to the new expectations. Patty described them as “anxiety relieving” as she had “almost felt like a failure last year” because she struggled to meet the

Friday postings on time. “To have that taken off my plate has been wonderful because I really didn’t ever understand why they had to be printed and by the door because they’re on a universal server that most people can access.” Similarly, Sandra shared that she previously,

was worried. And maybe I didn’t need to be. I was worried that if I felt like I had to change something due to what I was noticing in the classroom or because we had a snow day or whatever, that that was going to be frowned upon [...] I felt pressure. And I’m kind of a Type A person so I felt pressure on myself. I don’t know if [Laura] really put that pressure, but I do feel like I also just questioned, why do you need that?

Overall, at least for Sandra, Patty, and Lisa, the expectations within the lesson plan posting requirement, as opposed to writing the lesson plans themselves, were the main issue they faced. Under the new administration, however, it appears that new lesson plan requirements have shifted, making the process less tedious and anxiety-inducing.

Student Assessment Scores: BrightStar

The final primary finding regarding how evaluation and student assessment policies guided teachers’ conceptualizations of expertise pertains to the student assessment technology BrightStar. Overall, it is a widely accepted tool across all participants, yet there are differing opinions on how it should be used in the classroom and how it can most benefit teachers and students.

In the majority of the interviews, teachers pointed to the significant benefits of the BrightStar program. Patty shared that at the beginning of her time at Forest Elementary, she did not like the BrightStar program, yet after attending trainings with her principal, her perspective soon changed:

I only wish that I had known what I knew in January last September. [...] I’m sure there’s other programs out there, but for a teacher and for a group of teachers to be able to compare, compile and look at students and be able to possibly flex them for interventions, it’s a great way to measure. It changed the way I taught.

Similarly, Sandra shared that it was BrightStar, not any sort of evaluation tool, that impacted her teaching most:

Throughout the year, you know, you want to see those [BrightStar] numbers changing, those colors changing, that growth changing... and that is what changes my teaching. Because if I see that three students are not making progress, or if I see that five students are now at a first grade level, that is what I feel pushes me. And I see a lot of our conversations as a building in terms of: What are you doing? What’s working? What do you need help with? The conversations are rarely geared around our evaluation [...] they’re geared towards [BrightStar].

Teachers like Patty and Sandra see BrightStar as a necessary and integral part of their pedagogical decision making as well as their concepts of teacher expertise. It appeared that the program’s data guided instructional decisions more than any other resource or tool.

While most teachers noted the benefits of BrightStar, some also pointed out its potential drawbacks. Beth proposed that the power of BrightStar all depends on how it is used by the teacher: I think that you can skew the scores a little bit depending on ... so if you only teach to [BrightStar] and you only work with your kids on [BrightStar] lessons and you only prep them for what they’re going to see on [BrightStar], I think you might have

a little bit more skewed scores [...] But if you use the [BrightStar] data just to guide other types of instruction, I think that becomes a better gauge of where students are [...] it's not the be-all-end-all. You can look at other pieces as well. But I do think that it can be a good starting point for other types of lessons.

Although she still supports the use of BrightStar in the classroom, she described it as a “good starting point” for lesson planning, as opposed to a “be-all-end-all” evaluative tool.

Other teachers, such as Sophie, presented a much more critical view of the program. After beginning her response with “I like BrightStar”, Sophie shared concerns about program accuracy once students are in second or third grade, wondering if it still shows “where they're at as a learner”:

The kids that aren't burnt out on it seem to enjoy it a lot. It gives them... they're manipulating things and bubbles are popping in, aliens are singing and they're reading books. And I do like that for a station piece. Do I feel like we should push them every single week to be on it for 50 minutes for math and reading? I think that's a lot. I wouldn't throw it out. I like the data I get from it. I just don't want... I just feel like it shouldn't be that one thing.

On one hand, Sophie appreciates how BrightStar can be used as an enjoyable station for students to rotate through. Yet, she struggles with how BrightStar is expected to be implemented. She goes on to state:

I will say each year I see growth a little more, a little more. And it's pushing us and pushing us and pushing us. It's just that other piece of, 'Who is it... who is it leaving behind? Whose it beating down?' That's where I'm at. It's not like I don't want to do [BrightStar] and it's not that it hasn't made me a better teacher. It's just that... other part of, 'But is it worth that?'

Sophie's questions of “Who is it leaving behind?” and “Is it worth it?” shows a shift from the reflections seen with her colleagues. While they were generally comfortable with the use of BrightStar as a guide for instruction as long as it was not used solely as an evaluative tool, Sophie pushed this conversation even further, concerned that collecting BrightStar data comes at a cost to some of her students.

Notably, Sandra, Patty, and Lisa shared that expectations around BrightStar had significantly shifted after the changes in school and district leadership. Students no longer were required to reach a certain number of minutes on the platform each week and the district decreased the number of BrightStar benchmark tests required per school year from four to three. Sandra shared that although some of this shift was due to changing district leadership, an additional part was likely due to the new school leadership. “Our building principal last year [Laura] was the person who really kind of spearheaded [BrightStar] in the district [...] it was kind of her baby.” She noted that the previous year, “a lot of our PLC time was geared around [BrightStar] and how students were performing on assessments and lessons, and that's just not a focus this year.” Now, BrightStar is used significantly less, and other assessment tools provide by the state or district curriculum are relied upon to regularly assess student learning.

When reflecting on this shift away from BrightStar, Patty began by stating that she “doesn't really have feelings on it either way.” She still uses the program, but she's “using more of it as a resource in addition to other resources, whereas last year, it was the main thing that drove my instruction.” Lisa shared a similar sentiment, as she now uses a variety of assessments provided by the state and district curriculum, instead of solely relying on BrightStar. She was initially nervous

about this transition, asking herself, “if I don’t have [BrightStar], how will I know if my students are mastering skills? How will I know what they need? How will I know?” She soon, however, realized that shifting from one single digital tool to a variety of assessment tools and techniques was “tremendous.” Previously “everything hinged on that one assessment piece,” and this partly bothered Lisa because,

the thing is, it’s one or two days in the life of a child. There’s so many extenuating circumstances in those one or two days. So I think several data points over several times is so much more helpful than such a high stakes test.

Interestingly, this shift away from BrightStar also appeared to improve her relationship with her annual evaluation:

You want to know the honest truth? Honest truth is my [BrightStar] scores were my evaluations for the past two years [...] I had reached the point where whatever my [BrightStar] data said is what my evaluation was, and that was that.

Lisa shared information that was included in her evaluation from the previous year, and while the document itself included notes about her skills outside of BrightStar, she was left feeling as if the student data was the only thing that mattered in the evaluation meetings with her previous school leader. While Sandra, Patty, and Lisa had not yet gone through the formal evaluation process under the new administration at the time of the follow-up interviews, it appears that this shift away from BrightStar has impacted not only how teachers are measuring and understanding student learning, but also how teachers themselves are measured and understood as experts.

Discussion and Recommendations

The purpose of this study was to explore how teachers understood teacher expertise under ESSA. Overall, it appears that teachers at Forest Elementary do not necessarily exist within the exact onto-epistemic regime that Holloway (2021) previously identified. Instead, a plurality of understandings of expertise were shared. Additionally, open critique of the evaluation process was common among most of the teachers interviewed, often framed as something that either did not necessarily influence their teaching or actively hurt their ability to perform at their best. Within this diversity of thought, however, the power of student assessment scores remained a persistent theme, especially pertaining to the adaptive technology program BrightStar. Even when teachers uninhibitedly critiqued the evaluation protocol, they quickly shared how influential BrightStar was on their pedagogical decision making as well as their understanding of who “expert” teachers were. This heavy reliance upon BrightStar appeared to shift under new leadership, however, and teachers quickly adapted to other assessment tools. Some teachers showed initial concern, but once new possibilities were introduced, BrightStar lost some of its luster, no longer maintaining its ultimate power over understandings of students as learners and teachers as experts.

While Holloway’s (2021) exact onto-epistemic regime was not found in this study, there were elements of the five technologies of risk management identified across the teacher’s experiences and understandings of expertise. Surveillance was primarily conducted via the weekly lesson plan submission requirement, with some describing it as a waste of their time and primarily benefiting those “looking in” to their classrooms. Yet, most teachers did not take issue with the idea of making their lesson plans available for review. Instead, it was the procedures within requirement, such as having them posted by Friday afternoon, that impacted teachers’ perceptions of lesson plan surveillance the most. Additionally, numericization was sweepingly present, primarily through the

use of BrightStar for student assessment, teacher evaluation, and pedagogical decision making. There was a clear distinction made between using BrightStar for guiding instruction as opposed to evaluation, with most teachers finding it harmful if only used for evaluative purposes. Nevertheless, numbers were consistently used to determine teacher effectiveness through the quantification of students as learners, although this also appeared to be changing as the new school and district leadership were strategically adjusting their relationship with the platform. The final three technologies of risk management, normalizing judgments, examination, and discipline, were not clearly or consistently found in the data collected for this study. Although many teachers felt anxiety and stress from some of the evaluation practices used at Forest Elementary, I found little evidence that there were school- and system-level strategies that consistently normalized teacher-to-teacher comparison and high-stakes judgments/examinations.

In neoliberalized educational policy landscapes, individuals are deemed (un)successful based on simple numerical metrics which often overlook the complex realities faced by all who engage in the educational system, including educators, students, families, and communities (Germain, 2022). This trend still has a strong foothold in U.S. education, but there is preliminary evidence that there may be space to begin to reimagine, or at least readjust, how teacher expertise is understood and conceptualized under ESSA. Importantly, this study challenges us to consider how local educational leaders are or are not leveraging this flexibility, especially within states that have not mandated high-stakes teacher evaluation policies. For example, CPSD required school leaders to connect lesson plan submission to the district evaluation policy, and it appears that some of the requirements at Forest Elementary paradoxically made it more difficult for some teachers to plan effectively. Or consider the school's heavy reliance upon BrightStar. CPSD required BrightStar to be the primary measure of student learning and there is evidence that the principal's personal belief in BrightStar intensified how it was utilized in Forest Elementary. While technologies of risk management have become commonplace since the implementation of NCLB, it is important to avoid the normalization and uncritical acceptance of their implementation. In contexts where there may be localized teacher evaluation flexibility, schools and districts should continue to scrutinize their own chosen evaluation policies and practices. Which technologies are specifically being used? Who do they primarily benefit? What are the consequences of these technologies' processes? How do teachers "render [them]selves docile" (Foucault, 1977, p. 184) and what are the implications of this docility? Ultimately, within the ESSA era, we risk replicating the "regime of truth" as identified by Holloway (2021) if we do not continue to challenge the limiting evaluation practices that may linger from the NCLB-era. And potentially more urgently, we risk forgetting that new possibilities surrounding teacher expertise may emerge when normalized practices are disrupted.

Conclusion

To understand how policymakers, educational leaders, and advocates should approach these potentialities, future research should address some of the limitations found in this study. First, when exploring how teachers understand expertise, researchers should consider sampling based on grade level or subject area. Since this study used snowball sampling, there is a likelihood that teachers seen as having less expertise were not recommended, causing more experienced and skilled teachers to instead be interviewed. This may have impacted which technologies were identified within the Forest Elementary teacher evaluation policy since normalizing judgments, examination, and discipline may be more heavily utilized when evaluating teachers seen as less successful. Additionally, this study was conducted in a predominantly white working- and middle-class community. It would be important to consider how schools and teachers within and from Black, Brown, Asian, and

Indigenous communities are constructing teacher expertise within their situated contexts. Understandings of expertise within this study noticeably did not include concepts such as being aware of culturally sustaining practices (Paris, 2012) and overlooked issues such as race, gender, class, language, and nationality. Since teachers with historically marginalized identities more often resist oppressive practices regardless of their institutionalization (e.g., Souto-Manning & Cheruvu, 2016), exploring teacher expertise within a more racially and culturally diverse community would provide much needed perspectives in understanding teacher expertise today.

In the end, studying teacher expertise in the era of ESSA has the potential to open new possibilities for change. In contexts where the most high-stakes evaluation policies have not taken hold, the new federal flexibility should be seriously considered. While evaluation technologies are likely here to stay, district and local leaders may be able to carve a path toward more humane, holistic, and situationally sensitive understandings of expert teachers and the students they serve.

Acknowledgements

I would like to sincerely thank all who provided input and/or feedback on this manuscript, including my committee members, Lisa Dorner, Emily Crawford-Rossi, and Rebecca Scott, my PhD supervisor, Sarah Diem, and the journal reviewers. I would also thank the teachers and school leaders who took time out of their busy lives to participate in the conversations shared in this study. It is only with the help of a community that this paper could be written, revised, and published.

References

- Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, 37(2), 65-75. <https://doi.org/10.3102/0013189X08316420>
- Au, W. (2009). *Unequal by design*. Routledge. <https://doi.org/10.4324/9780203892046>
- Baker, B. D., Oluwole, J. O., & Green, P. C. (2013). The legal consequences of mandating high stakes decisions based on low quality information: Teacher evaluation in the Race-to-the-Top era. *Education Policy Analysis Archives*, 21(5). <https://doi.org/10.14507/epaa.v21n5.2013>
- Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, 18(2), 215–228. <https://doi.org/10.1080/0268093022000043065>
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35(5), 463-482. [https://doi.org/10.1016/S0883-0355\(02\)00004-6](https://doi.org/10.1016/S0883-0355(02)00004-6)
- Birman, B., Le Floch, K. C., Klekotka, A., Ludwig, M., Taylor, J., Walters, K., Wayne, A., Yoon, K., Vernez, G., Garet, M. S., & O'Day, J. (2007). *Evaluating teacher quality under No Child Left Behind*. Rand. <https://doi.org/10.7249/RB9287>
- Bleiberg, J., Brunner, E., Harbatkin, E., Kraft, M. A., & Springer, M. (2021). *The effect of teacher evaluation on achievement and attainment: Evidence from statewide reform* [Working paper]. The Annenberg Institute at Brown University. <https://www.edworkingpapers.com/ai21-496>
- Bradbury, A., & Roberts-Holmes, G. (2018). *The datafication of primary and early years education: Playing with numbers*. Routledge. <https://doi.org/10.4324/9781315279053>
- Brinkmann, S. (2014). Unstructured and semi-structured interviewing. In P. Leavy (Ed.), *The Oxford handbook of qualitative research* (pp. 277-299). Oxford Library of Psychology. <https://doi.org/10.1093/oxfordhb/9780199811755.013.030>
- Carterville Public School District (CPSD). (2017). *Supporting teaching and learning* [PowerPoint slides].
- Clarke, M. (2013). Terror/enjoyment: performativity, resistance and the teacher's psyche. *London Review of Education*, 11(3), 229-238. <https://doi.org/10.1080/14748460.2013.840983>

- Close, K. Amrein-Beardsley, A., & Collins, C. (2020). Putting teacher evaluation systems on the map: An overview of state's teacher evaluation systems post-Every Student Succeeds Act. *Education Policy Analysis Archives*, 28(58), 1-31. <https://doi.org/10.14507/epaa.28.5252>
- Cochran-Smith, M., & Lytle, S. L. (2006). Troubling images of teaching in *No Child Left Behind*. *Harvard Educational Review*, 76(4), 668-697. <https://doi.org/10.17763/haer.76.4.56v8881368215714>
- Daliri-Ngametua, R. (2022). Students' experiences of overt data-talk in the classroom: "It's all just this stupid system." In S. Riddle, A. Heffernan, & D. Bright (Eds.), *New perspective on education for democracy: Creative responses to local and global challenges* (pp. 94-109). Routledge. <https://doi.org/10.4324/9781003145806-10>
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1), 1-44. <https://doi.org/10.14507/epaa.v8n1.2000>
- Darling-Hammond, L., Bae, S., Cook-Harvey, C. M., Lam, L., Mercer, C., Podolsky, A., & Stosich, E. L. (2016, April). *Pathways to new accountability through the Every Student Succeeds Act*. Learning Policy Institute. <https://doi.org/10.54300/966.414>
- Ellingson, L.L., & Sotirin, P. (2020). *Making data in qualitative research: Engagements, ethics, and entanglements*. Routledge. <https://doi.org/10.4324/9780429058240>
- Englund, H., & Gerdin, J. (2019). Performative technologies and teacher subjectivities: A conceptual framework. *British Educational Research Journal*, 45(3), 502-517. <https://doi.org/10.1002/berj.3510>
- Finch, M. A. (2017). Tennessee to the top: One state's pursuit to win Race to the Top. *Educational Policy*, 31(4), 481-517. <https://doi.org/10.1177/0895904815604216>
- Ford, T. G., Van Sickle, M. E., Clark, L. V., Fazio-Brunson, M., & Schween, D. C. (2017). Teacher self-efficacy, professional commitment, and high-stakes teacher evaluation policy in Louisiana. *Educational Policy*, 31(2), 202-248. <https://doi.org/10.1177/0895904815586855>
- Foucault, M. (1977). *Discipline and punish: The birth of the prison* (A. Sheridan, Trans.). Vintage. (Original work published 1975).
- Germain, E. (2022). Well-being and equity: A multi-disciplinary framework for rethinking educational policy. *Peabody Journal of Education*, 97(1), 6-17. <https://doi.org/10.1080/0161956X.2022.2026716>
- Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. National Comprehensive Center for Teacher Quality. <https://files.eric.ed.gov/fulltext/ED521228.pdf>
- Guthrie, J. W., & Springer, M. G. (2004). *A Nation at Risk* revisited: Did "wrong" reasoning result in "right" results? At what cost? *Peabody Journal of Education*, 79(1), 7-35. https://doi.org/10.1207/s15327930pje7901_2
- Hallinger, P., Heck, R. H., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Educational Assessment, Evaluation and Accountability*, 26, 5-18. <https://doi.org/10.1007/s11092-013-9179-5>
- Hayes, W. (2004). *Are we still a nation at risk two decades later?* ScarecrowEducation.
- Holloway, J. (2021). *Metrics, standards, and alignment in teacher policy: Critiquing fundamentalism and imagining pluralism*. Springer Singapore. <https://doi.org/10.1007/978-981-33-4814-1>
- Holloway, J., & Brass, J. (2018). Making accountable teachers: The terrors and pleasures of performativity. *Journal of Education Policy*, 33(3), 361-382. <https://doi.org/10.1080/02680939.2017.1372636>

- Howell, W. G. (2015). Results of President Obama's Race to the Top: Win or lose, states enacted education reforms. *Education Next*, 15(4), 58-66. <https://www.educationnext.org/results-president-obama-race-to-the-top-reform/>
- Jeffrey, B. (2002). Performativity and primary teachers relations. *Journal of Education Policy*, 17(5), 531-546. <https://doi.org/10.1080/02680930210158302>
- Lavigne, A. L. (2014). Exploring the intended and unintended consequences of high-stakes teacher evaluation on schools, teachers, and students. *Teachers College Record*, 116(1), 1-29. <https://doi.org/10.1177/016146811411600103>
- Leitch, V. B., Cain, W. E., Finke, L. A., McGowan, J., Sharpley-Whiting, T. D., & Williams, J. J. (Eds.). (2018). *The Norton anthology of theory and criticism* (3rd ed.). W. W. Norton.
- Levinson, B. A. U., Sutton, M., & Winstead, T. (2009). Education policy as a practice of power: Theoretical tools, ethnographic methods, democratic options. *Educational Policy*, 23(6), 767-795. <https://doi.org/10.1177/0895904808320676>
- Lincoln, Y. S. (2015). Critical qualitative research in the 21st century: Challenges of new technologies and the special problem of ethics. In G. S. Cannella, M. S. Pérez, & P. P. Pasque (Eds.), *Critical qualitative inquiry: Foundations and futures* (pp. 197-214). Left Coast Press.
- Mehta, J. (2015). Escaping the shadow: *A Nation at Risk* and its far-reaching influence. *American Educator*, 39(2), 20-26. <https://eric.ed.gov/?id=EJ1064157>
- Milner IV, H. R. (2013). *Policy reforms and de-professionalization of teaching*. National Education Policy Center. <https://nepc.colorado.edu/publication/policy-reforms-deprofessionalization>
- Missouri Department of Elementary and Secondary Education (DESE). (2021a). *Building demographic data* [Data set]. <https://apps.dese.mo.gov/MCDS/Home.aspx>
- National Center for Education Statistics (NCES). (2020). *Locale Lookup*. <https://nces.ed.gov/programs/maped/LocaleLookup/>
- National Commission on Excellence in Education (NCEE). (1983). *A nation at risk: The imperative for educational reform*. U.S. Department of Education.
- Palmer, D. J., Stough, L. M., Burdinski Jr., T K., & Gonzales, M. (2005). Identifying teacher expertise: An examination of researchers' decision making. *Educational Psychologist*, 40(1), 13-25. https://doi-org/10.1207/s15326985ep4001_2
- Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher*, 41(3), 93-97. <https://doi.org/10.3102/0013189X12441244>
- Rose, N. (1999). *Powers of freedom: Reframing political thought*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511488856>
- Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). SAGE.
- Savin-Baden, M., & Major, C. H. (2013). *Qualitative research: The essential guide to theory and practice*. Routledge.
- Souto-Manning, M., & Cheruvu, R. (2016). Challenging and appropriating discourses of power: Listening to and learning from early career early childhood teachers of color. *Equity & Excellence in Education*, 49(1), 9-26. <https://doi.org/10.1080/10665684.2015.1121793>
- Stebbins, R. A. (2001). *Exploratory research in the social sciences*. Sage. <https://doi.org/10.4135/9781412984249>
- Tieken, M. C. (2014). *Why rural schools matter*. University of North Carolina Press. <https://doi.org/10.5149/northcarolina/9781469618487.001.0001>
- Viruru, R. (2009). Postcolonial technologies of power: Standardized testing and representing diverse young children. *Counterpoints*, 369, 100-118. <https://www.jstor.org/stable/42980384>
- Viteritti, J. P. (2013). The federal role in school reform: Obama's Race to the Top. *Notre Dame Law Review*, 87(5), 2087-2122. <https://scholarship.law.nd.edu/ndlr/vol87/iss5/10>

Wilkins, C. (2011). Professionalism and the post-performative teacher: New teachers reflect on autonomy and accountability in the English school system. *Professional Development in Education*, 37(3), 389-409. <https://doi-org/10.1080/19415257.2010.514204>

About the Author

Madeline Good

University of Missouri

Mewcq8@umsystem.edu

<https://orcid.org/0000-0002-5225-3355>

Madeline Good is a former public school teacher and current doctoral candidate in the Department of Educational Leadership and Policy Analysis at the University of Missouri. Her interests include how policies interact with the sociological and political aspects of education, especially regarding teacher expertise and issues of equity.

education policy analysis archives

Volume 31 Number 55

May 16, 2023

ISSN 1068-2341



Readers are free to copy, display, distribute, and adapt this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, the changes are identified, and the same license applies to the derivative work. More details of this Creative Commons license are available at <https://creativecommons.org/licenses/by-sa/4.0/>. **EPAA** is published by the Mary Lou Fulton Teachers College at Arizona State University. Articles are indexed in CIRC (Clasificación Integrada de Revistas Científicas, Spain), DIALNET (Spain), [Directory of Open Access Journals](#), EBSCO Education Research Complete, ERIC, Education Full Text (H.W. Wilson), QUALIS A1 (Brazil), SCImago Journal Rank, SCOPUS, SOCOLAR (China).

About the Editorial Team: <https://epaa.asu.edu/ojs/index.php/epaa/about/editorialTeam>

Please send errata notes to Audrey Amrein-Beardsley at audrey.beardsley@asu.edu

Join EPAA's Facebook community at <https://www.facebook.com/EPAAAPE> and Twitter feed @epaa_aape.