education policy analysis archives
A peer-reviewed, independent, open access, multilingual journal

epaa aape

Arizona State University

Volume 27 Number 137 October 28, 2019 ISSN 1068-2341

De-Evolution of Expectations for Evidence-Based Practices in Public Education in the United States

Michael P. Brady
Florida Atlantic University
&
Michael Hazelkorn
K & M Coastal Consultants
United States

Citation: Brady, M. P., & Hazelkorn, M. (2019). De-evolution of expectations for evidence-based practices in public education in the United States. *Education Policy Analysis Archives, 27*(137). https://doi.org/10.14507/epaa.27.4590

Abstract: Public education in the United States has been undergoing a shift from an empirical tradition in which practices and policies are derived from research, practice, reflection, and implementation. In this empirical tradition, professionals embrace a culture and commitment to evidence-based practices (EBPs) and expect that practices and policies in the field are supported by rational, data-driven models. In this paper, we present an argument and three cases that illustrate how educators have been undergoing a gradual shift away from empiricism toward a de-evolution of EBP. We propose that this gradual shift is based on a political-social context, in which practices and policies are implemented using the language of an accountability model of reform, in which national and state regulations, and accreditation bodies, establish expectations often devoid of an empirical basis for the practices they mandate.
Keywords: School reform; Evidence-based practice; Unintended consequences

Journal website: http://epaa.asu.edu/ojs/
Facebook: /EPAAA
Twitter: @epaa_aape

Manuscript received: 3/21/2019
Revisions received: 5/26/2019
Accepted: 5/26/2019

Devolución de expectativas para prácticas basadas en evidencia en educación pública en los Estados Unidos
Resumen: La educación pública en los Estados Unidos ha experimentado un cambio desde una tradición empírica en la que las prácticas y políticas se derivan de la investigación, la práctica, la reflexión y la implementación. En esta tradición empírica, los profesionales adoptan una cultura y un compromiso con las prácticas basadas en evidencia (EBP) y esperan que las prácticas y políticas en el campo estén respaldadas por modelos racionales basados en datos. En este artículo, presentamos un argumento y tres casos que ilustran cómo los educadores han experimentado un cambio gradual desde el empirismo hacia el desarrollo de la PBE. Proponemos que este cambio gradual se base en un contexto político-social, en el cual las prácticas y políticas se implementan utilizando el lenguaje de un modelo de reforma de rendición de cuentas, en el cual las regulaciones nacionales y estatales, y los organismos de acreditación, establecen expectativas a menudo desprovistas de un enfoque empírico. base para las prácticas que ordenan.
Palabras-clave: reforma escolar; Práctica basada en la evidencia; Consecuencias involuntarias

Devolução de expectativas para práticas baseadas em evidências na educação pública nos Estados Unidos
Resumo: A educação pública nos Estados Unidos sofreu uma mudança de uma tradição empírica na qual práticas e políticas derivam de pesquisa, prática, reflexão e implementação. Nessa tradição empírica, os profissionais adotam uma cultura e comprometimento com as práticas baseadas em evidências (PBE) e esperam que as práticas e políticas no campo sejam apoiadas por modelos racionais baseados em dados. Neste artigo, apresentamos um argumento e três casos que ilustram como os educadores experimentaram uma mudança gradual do empirismo para o devolução do PBE. Propomos que essa mudança gradual seja baseada em um contexto político-social, no qual práticas e políticas são implementadas usando a linguagem de um modelo de reforma de responsabilização, no qual regulamentos nacionais e estaduais e organismos de acreditação, defina expectativas muitas vezes desprovidas de uma abordagem empírica. base para as práticas que eles ordenam.
Palavras-chave: reforma escolar; Prática baseada em evidências; Consequências não intencionais

De-Evolution of Expectations for Evidence-Based Practices in Public Education in the United States

In recent years, many educators have begun to publicly lament the extent to which educational policies and practices have been adopted with little research that demonstrates their effectiveness (Cochran-Smith, Piazza, & Power, 2013; Cook & Cook, 2013; Rueter & Simpson, 2012). Once considered the gold standard for educational practice, policies and regulations that govern practice were expected to have solid evidence of effectiveness (i.e., a practice would show an actual, positive impact on the teachers, children, or others who were the "targets" of the practice) before these practices could become commonplace. Indeed, for many years, the very absence of a rigorous commitment to a science of implementation in public education was a major cause for concern among educational researchers, funding agencies, and accreditation bodies (Brittingham, 2009; Cook, 2002; Greenberg, Putnam, & Walsh, 2014; Whitehurst, 2012). As public educators

increasingly resisted curriculum and instructional practices that lacked objective evidence of effectiveness but had growing commercial support, numerous disciplines within public education adopted educational reforms, self-regulation, and increased accreditation standards, and a cultural shift among educators evolved who insisted on implementing interventions only if they were informed by evidence of effectiveness (McGuire, 2009; Rueter & Simpson, 2012; Yell, Conroy, Katsiyannis, & Conroy, 2013).

Ironically, the recent movement to *abandon* certain evidence-based practices (EBPs) has emerged as the result of legislative and regulatory efforts to "strengthen and reform" public education. Educators have faced growing criticism from an increasing number of critics eager to couch their input in the language of "accountability" and "educational reform" (Greenberg et al., 2014). Frequently, these accountability calls accompany changing political climates and election cycles, and include proposals to increase the regulation of public education, with many solutions directly linked to large-scale commercial contracts with private vendors (Ball, 2018; Ballou & Springer, 2015; Cochran-Smith et al., 2013; Lincove, Osborne, Dillon, & Mills, 2014). Au and Ferrare (2015) described the creation of a commercially-driven ideology that creates new markets in public education, and restructures relationships between citizens and the role of government that oversees public education. This new "corporate educational reform" creates new opportunities for commercial interests to participate in the regulation and governance of public education, and has resulted in a host of market-driven initiatives that have been successful in accessing public (taxpayers') assets. These commercial interests include numerous high-stakes student and teacher assessment initiatives, for-profit teacher certification schemes, large scale publically funded curriculum development and adoptions, and various virtual learning initiatives that cross nearly all disciplines in public education (Au & Ferrare, 2015; Ball, 2018; Ravitch, 2013).

Calls for reform and increased accountability have been seen in nearly all disciplines in public education, including policies and practices for young children (Ledford et al., 2016), elementary and secondary students (Papay, 2011), students with disabilities (Detrich, Keyworth, & States, 2016), English Language Learners (Jones, Buzick, & Turkan, 2013), new teacher preparation (Lewis & Young, 2013), in-service professional development (Collins, 2014), and others (Berliner & Glass, 2014). These calls for increased accountability by critics of public education are not a unique American phenomenon. Indeed, educators in other western countries have cautioned that educational reforms are increasingly driven by changing political ideologies and by commercial interests (Ball, 2018; Head, 2016; Lingard, 2013; Rowe & Skourdoumbis, 2019). In the United States, however, this phenomenon has been integrated into substantial federal legislation, and the examples in this paper illustrate the U.S. experience.

A prime example of this phenomenon is seen in the *A Nation at Risk* report (National Commission on Excellence in Education, 1983), in which critics called for regulations to (a) change content and curriculum required in schools, (b) increase expectations for high school graduates, (c) create mandates for time spent learning, and (d) proscribe particular instructional methods to be implemented in classrooms. Each of these mandates served as the impetus for increases in standardized testing of children as a way to promote accountability (Jorgensen & Hoffmann, 2003). As a result, the commercial testing industry has gained a dramatic increase in business with states and school districts.

A Nation at Risk opened the doors to four sets of national legislation that fundamentally changed public education. Each set of legislation called for school reform and accountability measures, few of which were grounded in EBPs. Many of these regulations incorporated standardized testing of K-12 students as the primary metric for measuring local, state, and national progress. For example, in 2001, the No Child Left Behind Act (NCLB) [the reauthorization of the

Elementary and Secondary Education Act (ESEA)] established a link between standardized student testing and teacher preparation, with exclusive contracts awarded to private vendors to manage states' assessment results. Not surprisingly, student test scores following the NCLB did not live up to Congressional intent to bring every American child to grade level in reading and math, nor did students' scores show any consistent link to their teachers' actual classroom performance. As additional attention to student standardized assessments and associated costs increased (with no evidence that increased testing was improving students' learning or teachers' performance), teachers, parents, and students grew increasingly frustrated ("The 49th Annual PDK Poll," 2017); as a result, many states requested and received waivers from these new regulations (Goldhaber, 2008; Jorgensen & Hoffmann, 2003). Ironically, these waivers required states to adopt yet another set of untested practices – using K-12 students' standardized assessment results to assign grades to schools in hopes of identifying "high achieving schools," then building these test scores into the evaluation systems for teachers and principals (Ayers & Owen, 2012). In subsequent years, additional national legislation and initiatives (e.g., the Every Student Succeeds Act and the Race to the Top competitive grant program) would stretch further the expectation that policies and practices should have evidence of effectiveness by requiring the use of K-12 student standardized test results to evaluate teacher preparation programs in colleges and universities (Brady & Miller, 2018; Lincove et al., 2014).

Many critics of public education have long been vocal in their reproach of educational practices (Finn 2013; Hess 2001; Walsh & Hale, 2004). Many have advocated "reforms" that increase regulation of teachers and schools; others propose to regulate curriculum and instructional practices, and even replace public education programs with private, vendor-driven programs (Burke, 2016; Henry & Bastian, 2015). A common theme across many criticisms is that (current) educational practices are not robust, and that educational practices frequently lack demonstrations of effectiveness (Whitehurst, 2012). Unfortunately, many of the actual policies, practices, and reform efforts that have emerged in the past two decades use accountability measures that have not been tested, and their effectiveness has not been demonstrated for their proposed purposes, thus violating a commitment to EBPs (American Educational Research Association [AERA], 2015; American Statistical Association, 2014). Reform advocates who are critical of K-12 public education often have little expertise in public education. Moreover, advocates with little expertise in public education frequently are linked less to genuine accountability efforts than they are to political philosophies and efforts to reduce support for public education. The result is a regulatory agenda that promotes "reform" efforts untested with actual children and teachers (see AERA, 2015; Berliner & Glass, 2014; Cook, 2002; Lewis & Young, 2013). By contrast, most education reformers with expertise *within the disciplines* of public education advocate that schools improve by enacting policies and practices that have undergone rigorous, field-based evaluations or have established a high degree of "practice-based evidence" of effectiveness (i.e., have been established as EBPs) (Cook et al., 2015; Strain, 2018). Although some critics have proposed that randomized controlled trials (RCTs) are the primary way to establish EBPs (Cook, 2002; Whitehurst, 2012), a broad array of research and practice actions have been used across various educational disciplines to establish practices that are effective including non-experimental group research designs (Chwalisz, 2003; Flay et al., 2005), qualitative inquiry (Giangreco & Taylor, 2003), single case experimental strategies (Horner et al., 2005; Kratochwill et al., 2013), and a focus on evidence informed by practice (McKnight & Morgan, 2019; Strain, 2018). Initiatives that lack an empirical basis typically weaken school systems (Cook, 2002).

How Are We Abandoning Our Commitment to Evidence-Based Practice?

If indeed the various disciplines in public education are experiencing a de-evolution in the expectation that practice, policies, and regulations should be based on evidence of effectiveness, then a reasonable question is: *how did we get here?* We propose two explanations for this phenomenon. First, as presented in the opening of this paper, we believe that a number of practices that lack any evidence of effectiveness have become adopted by public educators through a constant, step-wise series of gradual political actions, resulting in regulatory mandates at the state and national level. Given the interconnected nature of political rhetoric, low voter participation, and many citizens' skepticism toward nuanced scientific explanation, the nature of many educational policies in the U.S. has become a "simple solution" amalgam of easy-to-mandate laws and regulations that are not always based on empirical evidence. As the untested policies described in the opening section (e.g., the use of K-12 student assessment results to determine state purchase of vendors' curriculum packages) are adopted without showing any evidence that the newly-implemented policies improve practice (e.g., increases in student learning, or decreases in teacher stress), it becomes easier for a skeptical citizenry to accept the *next* untested policy (e.g., apply those same test results to teachers' evaluation). Over time, it is easy to see how an educational practice with little to no evidence of effectiveness gets adopted, particularly when mandated by state and federal regulation, and promulgated by accreditation standards, with few educators stopping to challenge the efficacy of a practice that has now become commonplace. We believe this phenomenon does indeed explain several current educational practices with little demonstration of effectiveness that are disrupting public education today. Numerous educators, for example, propose that the use of K-12 student standardized assessments have become the primary data to evaluate teachers in exactly this manner (Cochran-Smith et al. 2013). In spite of minimal evidence that these student data provide a profile of teacher effectiveness, *including effectiveness profiles for teachers and teacher candidates who did not actually deliver instruction to the students prior to their assessments*, this practice of Value Added Modeling (VAM) is now considered an acceptable practice in most states and school districts (Berliner, 2013; Lavigne, 2014). Ironically, teacher and teacher candidate evaluations that incorporate actual lessons delivered to K-12 students by these individuals have not become a common-place VAM alternative (Brady, 2019).

A second explanation might exist for some of the historical examples of ineffective practices in public education during the last 50 years. Educational researchers and practitioners have much to be proud of in discovering and promoting robust and powerful interventions and practices (see, for example the history of early intervention for young children with disabilities; structured programs such as Direct Instruction; and various configurations of meta-cognitive learning strategy instruction). However, given the propensity of educators to explore and experiment *on-the-run*, public education has also been plagued with some embarrassing mistakes during the past 50 years. After World War II and the launch of Sputnik, with a sense that public education in the U.S. was not what it should be, pedagogical practices like "New Math" (e.g., Beberman, 1962; Begle, 1968) and Open Classrooms (Perrone, 1972; Silberman, 1971) became commonplace during the 1960s and 1970s. With good intentions, educators implemented an era of "New Math" in which children might learn mathematics relations without studying math skills and operations (few learned either!) (Bond, 2005: Kline, 1973; Miller, 1990; Vigdor, 2013). Other educators implemented a generation of reading "instruction" virtually devoid of any instruction at all, as a means of promoting a love for reading, only to discover that few children would love to participate in activities that they could not perform. The Great Debate (Chall, 1967, Hempenstall, n.d.) about how best to teach children to read pitted those who believed in the skills-based instruction of phonics (Carnine & Silbert, 1979; Engelmann & Carnine, 1991; Flesch, 1955) against those who believed in a humanistic, holistic process of reading where children would learn to read naturally (Goodman, 1967; Smith, 2004). And of course, much

of this learning would occur in open classrooms, devoid of walls, so that children would be free to explore their spaces…. only to discover that their teachers quickly built walls to create parameters that would enable them to create physical spaces to help capture their children's attention and direct their focus to learning tasks (Cuban, 1984, 2004; Rothenberg, 1989). These practices (and others) had thoughtful advocates who invested logic and theory (and public dollars) into promoting reforms that they believed would change educational practice for the better. What they often lacked was evidence that the practices were effective, for certain students, under certain conditions (Bateman, 1991; Cromwell, 2016; Maddox & Feng, 2013; National Reading Panel, 2000; Stahl & Kuhn, 1995).

If these two explanations describe how public education is facing a reduction in the expectation that it's practices must maintain an evidence base of effectiveness, how is such an evidence base actually established in education? Public educators, like most professionals, have traditions and customs that constitute common practice, but not all common practices have evidence to establish them as effective. Many observers noted that educators have adopted expectations from medicine, often ignoring clinical evidence in favor of research methods that rely on narrow traditions (McKinght & Morgan, 2019). Evidence in medicine, and across many public agencies, often reflects a hierarchy of information gathering methods, where knowledge established by experimental research methods using RCTs, is considered the apex of evidence (Head, 2016; Horntvedt, Nordsteien, Fermann, & Severinsson, 2018). Such traditions, however, have numerous critics, who point out that experimental methods are only one class of procedures that generate actionable evidence, and who paint the evidence with a binary logic—either effective or not. Effective practices are better considered as being *informed* by evidence, and this evidence should include knowledge gained through implementation and delivery of services (Head, 2016; Lingard, 2013; McCall & Green, 2004; Strain, 2018).

Within public education, standards that establish EBPs vary dramatically across disciplines and research methods. EBPs in education are "practices that are supported by multiple, high-quality studies that utilize research designs from which causality can be inferred and that demonstrate meaningful effects on student outcomes" (Cook & Cook, 2013, p. 73). Yet, defining and deciding which practices are evidence-based, is difficult due to the variety and complexity of disciplines in education (Lancaster & Bain, 2019; Lewis, Hudson, Richter, & Johnson, 2004; Odom et al., 2005; Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008). Adding to this complexity is variability in research designs and methods regularly used by educational researchers to investigate and validate interventions for different types and ages of learners and curriculum content. In many educational disciplines, these designs and methods include *group designs* grounded in educational and psychological research methods, as well as experimental *single subject designs* grounded in the traditions of behavioral psychology. And, as Strain (2018) reminds us, evidence generated through elegant experimentation often bear little resemblance to actual implementation and service delivery, further supporting the need for practice-based evidence as a measure of face validity in education.

Standards used by intervention researchers using *group research designs* typically include evidence that results are (a) generated using rule-governed methods; (b) obtained from at least two rigorous trials; and (c) presented with clear participant, measurement, and analysis descriptions. The evidence (a) demonstrates consistent effects, (b) is observed and verified independently, (c) includes at least one example of long-term effects, and (d) includes detailed replication information from other researchers. The effects are interpreted in relation to "proof and rationality" and are generated using a hierarchical design approach ranging from (a) random assignment, (b) clinical trials in which some aspects of the most rigorous standards might be missing (but not "fatal flaws"), (c) case control studies based on retrospective treatment data, (d) secondary data analysis including meta-analyses, (e) "impressionistic" reviews not based on secondary data analysis, and finally to (f) case

studies and other reports without rigorous methods (Chwalisz, 2003; Flay et al., 2005; Rycroft-Malone et al., 2004). Although many educational researchers who use group designs argue that RCTs are the only designs that meet an objective standard for establishing the effectiveness of interventions, the hierarchy of group research designs described here are frequently advocated as a more rational approach (McCall & Green, 2004).

Standards employed by intervention researchers using *single subject research design methods* (Horner et al., 2005; Kratochwill et al., 2013) typically include evidence that (a) designs show a causal relationship between the independent and dependent variables; (b) within- and between-subject comparisons exist; (c) controls are established for major validity threats and to allow for systematic replication; (d) dependent variables are selected for their social significance and are measured repeatedly, by at least two observers, and the degree of agreement of their observations is reported; (e) independent variables under investigation are evaluated for fidelity; and (f) baseline conditions serve as a comparison condition for measuring the dependent variable. Further, when using single case research designs, each participant serves as his or her own unit of analysis with (typically) 3-8 participants per study. Participants, settings, variables, and the process for selecting these elements must be operationally defined.

With such rigorous standards established by disciplines and researchers, how do practices and policies become established for educators? First, many practices and policies become established because they are (a) grounded in empirical evidence, (b) effective, and thus (c) adopted by teachers and principals. These practices are taught in preservice classes for future teachers, and promoted during inservice professional development activities for current teachers. Second, other practices and policies become established because they are *mandated as regulatory measures*, *regardless of whether they are based on evidence that they are effective*. The intent of much of the national legislation has been to promote effective practices. For example, when the NCLB became law, it emphasized accountability, and mandated that schools use programs and practices based on scientifically-based research. Similarly, the re-authorization of the Individuals with Disabilities Education Act in 2004 required schools to use EBPs "to the extent practicable" to improve student outcomes, and the Every Student Succeeds Act in 2015 required that teachers use EBPs to teach high academic standards to all students to prepare them to succeed in college and careers. However, these regulatory efforts also established policies and practices that have little empirical support for their effectiveness. And last, accreditation agencies play a role in advancing many policies and practices, regardless of the efficacy of their effectiveness.

Because accreditation in the United States can be traced back to the 1880s and 1890s when the New England, Middle States, North Central, and the Southern associations of colleges and schools were founded and established minimum standards for their institutions (Brittingham, 2009), accrediting agencies have become significant players in promoting EBPs, as well as other policies mandated by national legislation. Accreditation agencies bridge the gap between teacher preparation programs and K-12 schools by establishing standards that both must meet (Achieve, 2009; Alliance for Excellent Education, 2014; Eaton, 2015; Garfolo & L'Huillier, 2015). Accreditation standards help assure that teacher preparation programs will graduate teachers who understand, select, and use EBPs to make an educational impact upon their students (Scheeler, Budin, & Markelz, 2016; Simonsen et al., 2008). Various efforts have been advocated to strengthen accrediting agencies' roles in promoting EBPs for teachers and teacher preparation programs, including (a) adopting particular practices, (b) merging agencies, (c) refocusing teaching standards on subject matter content, and (d) integrating teacher preparation into deeper clinical experiences (Alter & Naiditch, 2012; Darling-Hammond, 2010). The effort to establish an evidence base for policies and practices is wasted, however, if teachers, administrators, and other practitioners are not familiar with these practices or

lack the professional judgement on when and how to implement them (Cook & Cook, 2013; Rueter & Simpson, 2012). If policies require certain practices, but educators do not know which practices are actually informed by evidence, or do not how to implement them with fidelity, then the policies will have "little impact or negative unintended consequences that harm the educational process" (Detrich et al., 2016, p. 129).

In the remainder of this paper, we provide three cases as examples where educators have had to adopt practices or policies with little to no empirical evidence of effectiveness in response to "accountability reforms" implemented by national and state regulations, and accreditation expectations over the past three decades. In lieu of a continued commitment to an empirical model in which research, practice, reflection, and implementation form the basis of educational policies and practices, our observation is that these cases demonstrate the de-evolution of a commitment to evidence-based practices in public education. An empirical model of decision-making is being replaced by a model based on a political-social context, with policies and practices being implemented based on political ideologies and regulatory contingencies, with little regard for their impact on students, teachers, families, or educators. As such, we suggest that these three cases involving (a) the effectiveness of teachers, (b) the evidence for student learning gains, and (c) the effectiveness of teacher preparation programs represent a move to abandon a commitment to an empirical basis for the field in a *de-evolution of EBPs*. The cases are summarized in Table 1.

Table 1
Examples of Evidence-Based Practices that are Being Abandoned

Practice	Evidence	Replaced By
Demonstration of teachers' classroom effectiveness	Observation of teachers' lesson delivery; review of K-12 student learning products.	Teacher performance on standardized assessments that do not include lesson planning, organization, delivery, or management; Entry into the teaching profession contingent on performance on standardized exams rather than demonstrated teaching performance.
Demonstration of K-12 student learning	Measurable, observable learning gains on discrete learning tasks of individual students.	Student learning limited to group results on standardized assessments, frequently on content not reflected in classroom curriculum.
Demonstration of teacher preparation program effectiveness	Observation of teacher candidates' lesson delivery in multiple practicum and student teaching placements; Review of K-12 student learning products; Graduates preparation to teach prior to accepting employment.	Certification of student teachers based on their knowledge performance on standardized assessments; Accreditation of Teacher Preparation Programs based on K-12 students' performance on standardized exams several years after being taught by the TPPs' student teachers.

Case #1: Abandoning Evidence of Teacher Effectiveness

Nearly every state in the US is experiencing a teacher shortage in some area of the educator workforce. In some states, the teacher shortage is most common in geographic areas (e.g., in rural areas or inner cities); elsewhere, the shortage is pervasive in specialization areas (e.g., STEM, special education, English Language Learners, or reading). States have responded to these shortages by pursuing a variety of initiatives, including "grow your own" teacher programs, alternative certification programs, and other initiatives.

As the number and variety of alternatives to increase the teacher pool has grown, the variability in the quality of the professional development efforts that accompany these programs also has expanded. While some of the alternatives have built carefully sequenced knowledge and skill development into their programs (Brownell, Rosenberg, Sindelar, & Smith, 2004; Darling-Hammond & Youngs, 2002), other programs included no professional development requirements at all (Keller, Brady, Duffy, Forgan, & Leach, 2008). Ironically, efforts to fill teacher vacancies with minimally trained teachers have created new dilemmas in many schools as the continuing teacher shortages have created increases in minimally effective teachers.

With growing public concern directed toward teachers who were unable to demonstrate their basic effectiveness in teaching children, increased accountability demands have been increasingly aimed toward teachers. And like the accountability logic applied to children, the accountability measures adopted for teachers have been standardized test measures that are easily administered, easily evaluated, and easily contracted to private vendors – measures that have virtually nothing to do with teachers' actual daily performance.

Researchers who study teacher effectiveness can draw on over four decades of evidence that demonstrates that teachers who deliver high quality instruction can improve the learning gains in the students whom they teach (Creemers & Reezigt, 1996; Darling-Hammond & Youngs, 2002; Grossman & McDonald, 2008; Jones et al., 2013). Further, this instruction can be reliably evaluated using a variety of measures (Berliner, 2005; Buzick & Laitusis, 2010; Cochran-Smith et al., 2013; Goldring et al., 2015; Swank, Taylor, Brady, & Freiberg, 1989). Ironically, the accountability measure currently being selected in many states as the preferred standard for determining whether teachers will become certified and remain a part of the teacher work force meets none of this empirically-based evidence and is at the heart of a new crisis in the development and retention of effective teachers.

Florida's Specific Case: Removing the Wrong Teachers Based on the Wrong Evidence

Florida is one example of a state that has had a teacher shortage for several years. Florida employed approximately 175,000 certified teachers for its K-12 schools in 2017-18, and the need for new teachers increases each year (Florida Department of Education, 2018). In 2016, there were 2,400 open teacher jobs; the number climbed to approximately 3,000 in 2017, and to over 4,000 by the beginning of the 2018-19 school year. To fill these teaching positions, Florida—like many states—participates in a large number of alternative teacher certification programs (Keller et al., 2008). Ironically, in 2015, Florida awarded a $57 million contract to Pearson Vue to redesign the standardized teacher certification examinations and raise the examination cut scores after concerns that the entry certification standards were too low. (In Florida, prospective teachers must pass standardized examinations to demonstrate their fluency in basic academic skills prior to admission to university-based teacher preparation programs. Prior to earning permanent teaching certification, teachers also must pass *subject specialty* examinations to demonstrate their subject matter expertise [e.g., English or Social Studies content], as well as *professional* examinations to demonstrate their

knowledge of issues facing the profession [e.g., communication with parents, curriculum standards, educational regulations]. The certification examinations consist of multiple choice and short answer test items that assess content acquisition.)

Did the redesigned exams make a difference? Indeed, they did; first time pass rates among teacher test takers *dropped* to 57%. As Florida prepared for the 2018-19 school year, state education officials learned that districts were forced to fire over 1,000 teachers they had previously identified as effective based on their classroom performance with children, due to these teachers' inability to pass the newly redesigned standardized examinations ("Demanding Answers," n.d.; LaGrone & Apthorp, 2017). Ironically, none of the certification examinations taken by these new teachers included any demonstration of teaching effectiveness, or any other evidence of teachers' direct impact on student learning. Amazingly, in a year when state officials estimated they would need 4,000 additional teachers for Florida's classrooms, district officials were forced to fire 1,000 existing teachers they reclassified as ineffective *based on new standards and standardized examinations that provide no evidence identifying the teachers as either effective or ineffective.*

Like other policy initiatives, abandoning evidence-based measures for hiring and certifying teachers likely has unintended negative consequences beyond exacerbating the immediate teacher shortage (Lavigne, 2014). The obvious impact of mis-identifying teachers as ineffective based on measures that do not identify teacher effectiveness does nothing to reduce the teacher shortage. Such a practice appears, instead, a high-probability practice for increasing professional dissatisfaction, and driving serious future educators away from K-12 classrooms.

Case #2: Abandoning Evidence of Student Learning

One measure of effective teaching includes the ability to assess students' classroom performance, and then adjust one's delivery of instruction based on that performance (Abbott & Wren, 2016; Al Otaiba et al., 2011). Strong teachers adjust the pace of instruction, provide remediation when needed, and deliver supplemental opportunities for independent learning. These decisions require that teachers are adept at assessing student learning in both overt and subtle ways, and can use that information to drive their instruction (Ardoin, Witt, Connell, & Koenig, 2005; Simonsen et al., 2008).

Unfortunately, many teachers enter the workforce from careers paths that did not prepare them to teach. Without performance assessment skills needed to influence their daily teaching decisions, these new teachers face school and district expectations that classroom assessment is synonymous with high-stakes standardized testing, mandated for accountability purposes. Without a background that teaches them the links between performance assessment, curriculum, and the teaching methods used in their classrooms, these teachers are socialized to "know" assessment as the regulatory mandates that increase the demands for standardized student testing as a way to promote accountability (Jorgensen & Hoffmann, 2003). But as Berliner (2017) reminds us, regulations to increase standardized testing do little to influence classroom learning.

A shift from student performance assessment that promotes K-12 student learning to one that promotes accountability was explicit in the NCLB Act. The NCLB Act firmly established the expectation that students' standardized assessment results would be the metric used (a) to identify high and low achieving schools, (b) as part of teacher and principal evaluation systems, (c) to make personnel decisions within schools and districts, and (d) to make funding and resource decisions within states and districts (Ayers & Owen, 2012; Jorgensen & Hoffmann, 2003). K-12 student assessment results took on a greater role with the next federal initiatives, the Race to the Top, and the Every Student Succeeds Act. Ironically, the mandates that link single, standardized measures of

K-12 student learning to these various policy decisions have little grounding in empirical evidence. In some cases, an evidence base is completely absent.

Student assessment is a fundamental part of the learning cycle before, during, and after instruction. Careful assessment of learning requires an enormous investment of teachers' time and energy, and yields extraordinary benefits to teachers and students. Teachers gain by using student assessment results to increase the efficiency of their instruction, and students benefit by receiving instruction that matches their readiness for learning. Indeed, performance assessment has been identified as an EBP for over four decades (Abbott & Wren, 2016; Deno, Fuchs, Marston, & Shin, 2001; Fuchs & Deshler, 2007). However, the evidence base that student assessment is a best, evidence-based practice for establishing student learning is met only when:

1. K-12 student performance is linked to curriculum-based measures of the content that students are actually being taught;
2. Student performance is assessed frequently enough to determine whether instruction has an effect on student learning;
3. Assessment results enable instructional feedback that acknowledge or improve student performance;
4. Assessment results enable teachers to make decisions and change their instruction to help students improve their performance;
5. Student assessments include a variety of genuine work samples that tap the range of student performance (Abbott & Wren, 2016; Brady & Miller, 2018; Deno et al., 2001).

An enormous investment in time and fiscal resources has become a requirement in state and national regulation and policy with no empirical support that linking standardized K-12 student assessments (a) would actually improve K-12 student learning, or (b) is an effective intervention for the many personnel, funding, or other policy decisions currently being linked to these standardized measures of student learning. That is, although student performance on high-stakes, standardized assessments fails to meet the standards as an EBP for *student learning*, it has now become the accepted standard for numerous national policy practices.

The Specific Cases: Accepting the Wrong Evidence as Evidence of Student Learning

In spite of repeated evidence that teachers who use student performance data improve learning outcomes and efficiency in their students (Al Otaiba et al., 2011; Ardoin et al., 2005; August, Francis, Hsu & Snow, 2006; Fuchs & Deshler, 2007), the default measure of student learning has become student performance on high-stakes, standardized assessments. Often criticized as unreliable for many populations of children (Holdheide, Goe, Croft, & Reschly, 2010; Jones et al., 2013; Steinbrecher, Selig, Cosbey, & Thorstensen, 2014), research over that last 40 years demonstrates that these assessments often produce data that misrepresent evidence of actual learning.

Research by Holdheide et al. (2010) demonstrated that students with disabilities, English Language Learners, and K-12 students who are not fluent readers often under-perform on standardized assessments, even when they performed at a mastery level on genuine academic classroom assignments with complex subject matter. Others have found that some sub-groups of students perform in an unreliable manner on standardized assessments over time (McCaffrey, Sass, Lockwood, & Mihaly, 2009) and across multiple assessments of the same skills (Papay, 2011). In contrast to standardized assessments, performance on authentic classroom measures are less likely to show this variance (Steinbrecher et al., 2014). Using the "wrong evidence" to understand student

learning has many of the same unintended consequences seen in other policy initiatives (Brady, Duffy, Hazelkorn, & Bucholz, 2014; Lavigne, 2014). Using standardized assessments of children to make decisions on practices that the tests were never intended to measure (and certainly on outcomes that were not part of the effort to norm the tests) diverts instructional time and effort from students, and creates a false impression of a scientific basis for policy decisions that have little merit in educational evidence. Far from a "do no harm" rationale, the impact of this practice has unintended negative consequences that affect students who traditionally do not show learning gains on standardized assessments (e.g., children in high poverty areas, low achieving students, children with disabilities), even when they perform well on classroom measures of learning.

Case #3: Abandoning Evidence of Effective Teacher Preparation Programs

Teacher preparation practices have changed frequently and dramatically throughout American history (Schneider, 2018). Once a random and parochial undertaking, current practices are often characterized as organized, even bureaucratic, with scaffolded professional development that incorporates knowledge-based experiences, clinical applications, and professional socialization (such as mentoring and support networks) (Darling-Hammond, 2010). Far from a uniform "industry" of teacher preparation, a host of alternatives exist that connect pedagogy to practice, often incorporating apprenticeships, residencies, various on-the-job training opportunities, alternative career path development options, as well as traditional university-based teacher preparation programs (TPPs) (Darling-Hammond, 2010).

In spite of the variety of TPP models, the same ideological critics of public education noted previously have also advocated teacher preparation "reforms" that increase regulation and oversight over the professionals and programs who prepare teachers, including recommendations to replace university-based TPPs with private, vendor-driven models (Kronholz, 2012; Walsh & Hale, 2004). Like other regulatory mandates, many of these educational initiatives are untested, and lack an empirical basis grounded in evidence of effectiveness.

For example, when the NCLB Act and the Race to the Top competitive grant program required that states link teacher evaluations to K-12 student assessments, there was no evidence that this link would actually improve K-12 student learning. When this link was extended further to evaluate the efficacy of the TPPs from which these teachers graduated (Kronholz, 2012), the policy became an example of using data for efforts *four times removed* from their original intent (Brady & Miller, 2018). That is, data generated from high-stakes, standardized K-12 student assessments, under scrutiny for their history of validity and reliability challenges for children (first-order decision-making) were being used to make decisions about TPPs. In effect, K-12 student data were being used to evaluate teachers, and then these teachers' schools, who in turn, were graduates of TPPs.

There are numerous flaws in the logic and evidence that K-12 students' performance on high-stakes assessments has a causal link to TPPs. First, the logic of this accountability link assumes that university-based TPPs are a controlling factor in the day-to-day teaching behavior of their graduates, years after they complete their degrees and training experiences. Second, the link also assumes that these teacher graduates have been the controlling factor in the learning of the K-12 students whom they teach as reflected on the standardized assessments. Neither assumption is accurate. Floden (2012), Berliner (2014), and others note the myriad factors that influence teachers' instructional performance. Although TPPs greatly influence how teachers perform in their initial teaching roles, other personal (family, economic, and health issues) and labor market factors (acquiring advanced certifications, employment mobility) affect teachers' performance after their initial employment. And the empirical literature is replete with evidence of the micro- and macro-

school and classroom factors that contribute to student learning over which teachers have little control (e.g., student attendance, class size, administrative policies, students' prior knowledge) (Berliner, 2005; Cochran-Smith et al., 2013; Creemers & Reezigt, 1996). Combined, these factors make absurd the assumption that TPPs can be evaluated meaningfully based on high-stakes K-12 student assessment data. At best, policies that mandate using K-12 student assessment data as the metric to evaluate TPPs lack an empirical basis needed to establish this as an EBP.

A Specific Case: Accepting Better Evidence of Teacher Preparation Effectiveness

For many years, teacher educators have advocated that TPPs have a responsibility to demonstrate that their teacher candidates and graduates make a positive impact on the instructional growth of the K-12 students whom they teach (Greenwood & Maheady, 1997; Shores, 1979). Although this is by no means a universally held position, it is also a position that acknowledges that numerous factors intervene between teachers, students, teacher candidates, and TPPs. However, many TPPs incorporate assignments into their programs that directly measure their candidates' impact on student learning during candidates' internship experiences.

In 2012, researchers at Florida Atlantic University (FAU) initiated a series of investigations to evaluate the impact of teacher candidates' use of curriculum-based measures on K-12 students' learning gains during various clinical experiences (student teaching, practicum, and graduate internships; Brady, 2019). Unlike a value-added model (VAM) based on high-stakes K-12 student assessments, the curriculum-based VAM at FAU examined whether lessons delivered by teacher candidates would result in learning gains in the students taught by these candidates. Each lesson delivered by a teacher candidate was aligned to a specific curriculum standard; candidates divided the curriculum content into smaller clusters of teachable units, developed learning objectives to match the content, and then delivered lessons designed to help students meet the objectives. Two sources of evidence were collected to measure whether the K-12 students showed learning gains. First, the teacher candidates collected pre-test information prior to their instruction, and post-test information several weeks after the lessons to determine whether students made any gains in learning as a result of the candidates' instruction. Second, after the clinical experience, university supervisors collated instruction on the percentage of K-12 students who met the learning objective established by their teacher candidates. Together, these two direct measures of learning provided evidence of the impact of the candidates' instruction on student learning – the very essence of a VAM accountability model. As important, these data provided information that the TPP used to improve teacher candidates' performance and to make program and curriculum improvements in the TPP.

In the first exploration (Brady, Heiser, McCormick, & Forgan, 2016), investigators standardized the protocol for evaluating K-12 students and teacher candidates. Undergraduate candidates in both student teaching and part-time practicum placements, and graduate students in their internships, showed a substantial impact on K-12 student learning, with students averaging 35 to 40 percentage points on pre-to-post learning gains in lessons delivered by the teacher candidate cohorts. Between 91-96% of K -12 students met their learning objectives. These student learning gains were statistically significant, with strong effect sizes. In a second curriculum-based VAM exploration, Brady, Miller, McCormick, and Heiser (2018) investigated whether teacher candidates might deliver more effective instruction as they progressed from part-time practicum to full-time student teaching. Again, K -12 students showed statistically significant pre-to-post-test changes in their learning, and a statistically significant number of K-12 students also met their learning objectives. As candidates progressed *from* their part-time practicum *to* their full-time student teaching, K-12 student learning gains continued to increase, and candidates' instruction had a positive impact on student learning, regardless of whether the nature of their instruction was purely

academic (e.g., solving math word problems), or the lessons were more practical in nature (e.g., planning community mobility excursions). The results of these two initial explorations validated the curriculum-based measures of student learning as an alternative to previous VAMs based on high-stakes K–12 student standardized assessments.

In a third study designed to explore whether specific teacher candidate behaviors might *predict* student learning gains (McCormick, Brady, Morris, Heiser, & Miller, 2019), investigators selected items from a classroom observation instrument that might predict whether K-12 students would meet the learning objectives on lessons delivered by the teacher candidates. Only observation indicators related to classroom management were reliable in predicting K-12 student learning gains. Although strong classroom management and organization skills have long been identified as an EBP among effective teachers, this link has not been integrated into the other VAM research to date. Finally, in a fourth curriculum-based VAM exploration, McCormick, Brady, Miller, Heiser, and Morris (2018) found that 5 years of undergraduate student teacher data showed their instructional impact on students' pre-to-post test scores, as well as on the numbers of students who met their specific learning objectives, was statistically significant with moderate to strong effect sizes. In addition, the correlation between these two measures of student learning was statistically significant, with a strong effect size. Several lesson delivery behaviors from the observation instrument were significantly related to students who met their learning objectives. Across these studies, the evidence was convincing. Using a curriculum-based VAM enabled the candidates to show their effectiveness, and provided useful feedback to the K-12 students, teacher candidates, and the TPP for program improvement (Brady, 2019).

Conclusion

In recent years, many educational policies and practices have been mandated that have little research or practice evidence to support their effectiveness (e.g., using K-12 student standardized assessments to evaluate teachers, and replacing public education programs with private, vendor-driven programs). Under the guise of greater accountability in public education, proponents of educational reform, including national and state legislators and accreditation agencies, have pushed for these new policies and practices. Many of these policies and practices have been advanced by advocates who often have little expertise in public education and are not responsible for implementing them (Berliner & Glass, 2014). Rather than advancing practices that are informed by evidence as promoting quality outcomes for students, these policies and practices have led to a de-evolution of expectations for EBPs in public education. As Brady et al. (2014) observed, these policy changes and practices risked "unintended effects never envisioned by the people who initiated the changes" (p. 102).

The de-evolution of expectations for EBPs across the various disciplines in public education suggests a future with many unanswered questions. For example, will the professional relationships among teacher educators, mentor teachers, and preservice teachers change? Will teachers become less collaborative, creating a culture contrary to that which is necessary for system reform (Fullan, 2011)? As the proportion of certified teachers who lack formal teacher preparation increases, will the value of certification as an indicator of teacher competence actually decrease (Brady et al., 2014)? As K-12 student assessment results play a larger role in evaluating TPPs, will these TPPs steer their candidates and graduates away from (or toward) certain school districts to obtain better scores for their programs—that is, will TPPs learn to "shop" for *high performing, low need* schools as their partners? Will in-service teachers be less amenable to mentoring preservice teachers because of the effect the preservice teachers may have upon student test scores? Are there disincentives for

teaching certain populations of students (e.g., gifted students, English Language Learners, students with disabilities) who are less likely to show large gains in annual test scores? Will teachers become distrustful, shy away from teaching these students (Brady et al., 2014), "game the system," or use other unprofessional or unethical behaviors to hand-pick the students whom they wish to teach (Collins, 2014)?

All three of the cases presented in this paper are linked to VAMs. Will the new VAMs, unsupported with evidence of effectiveness, be used "to identify the lowest performing kids to pull out for tutoring or remediation, and also the 'bubble kids' [upon] whom…[to focus] their teaching efforts…to try to maximize growth scores" (Collins, 2014, p. 14)? Or, as Berliner (2013) described, will teachers realize that certain students are "money kids" and select those students to be in their classrooms because they would expect them to score well on the standardized assessments, show the most growth during the year, and therefore enhance the possibility of stronger evaluations and bonuses for the teachers? Conversely, what will future schools look like if principals become less willing to house certain high-risk, special programs in their schools *because* students in these schools are likely to lower their schools' ratings (Brady et al., 2014)? Aside from lacking evidence of effectiveness, the unintended consequences of these policies would be unfortunate indeed.

As Collins (2014) found, with an obsession with test scores and a subsequent drive for educators' to teach to the test, we may see a substantial decrease in real teaching and little real student learning. Students might well become "less likely to think and inquire and innovate, and more likely to sit-and-get. Raising a generation of children under these circumstances seems best suited for a country of followers, not inventors, not world leaders" (p. 18). One can imagine the low morale among teachers who don't understand how they are being evaluated, or how they might use the results of VAMs to improve their teaching. These are clearly not the conditions for retaining the best and brightest for public education and teaching as a profession.

Defining and deciding which practices are informed by evidence, and preparing teachers to implement these practices, is difficult at best. Unfortunately, this task has become more difficult as individuals and organizations with political and commercial interests in the practices gain greater roles in educational governance. A further de-evolution of EBPs, with little to no efficacy data to support the effectiveness of policies and practices for decision-making is more than troubling. Using data for purposes for which they were never designed (such as high-stakes testing to establish VAMs), and then evaluating the effectiveness of teachers, student learning gains, and teacher preparation programs will ultimately affect teaching as a profession and does not bode well for the future of public education. However, educational policies and practices informed by evidence of effectiveness might have intended positive consequences that do not result in a de-evolution of evidence-based practice. Indeed, they might even result in anticipated *positive* changes in students, teachers, administrators, and teacher educators.

Acknowledgements

The authors wish to thank several individuals for their input and suggestions during the development of this paper: Drs. Frank Di Vesta, Dennis McDougall, Lisa Finnegan, Mary Little, Barbara Ridener, and the participants of the Oxford University Round Table.

References

Abbott, A. L., & Wren, D. G. (2016). Using performance task data to improve instruction. *The Clearing House, 89*(1), 38-45. https://doi.org/10.1080/00098655.2016.1138924

Achieve, Inc. (2009). *Accelerating college and career readiness in states: Teacher effectiveness*. Washington, DC: Author. Retrieved from ERIC database. (ED508082)

Alliance for Excellent Education. (2014). *On the path to equity: Improving the effectiveness of beginning teachers*. Retrieved from the Alliance for Excellent Education website: https://all4ed.org/reports-factsheets/path-to-equity/

Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011). Assessment data-informed guidance to individualize kindergarten reading instruction: Findings from a cluster-randomized control field trial. *The Elementary School Journal*, *111*, 535-560. https://doi.org/10.1086/659031

Alter, M., & Naiditch, F. (2012). Teacher education at the crossroads: Burning questions that just won't go away. *Critical Questions in Education*, *3*, 65-82. Retrieved from https://academyforeducationalstudies.org/journals/journal/current-and-past-issues/200-2/

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

American Statistical Association. (2014). ASA statement on using value-added models for educational assessment. Retrieved from the ASA website: http://www.amstat.org/

Ardoin, S. P., Witt, J. C., Connell, J. E., & Koenig, J. L. (2005). Application of a three-tiered response to intervention model for instructional planning, decision making, and the identification of children in need of services. *Journal of Psychoeducational Assessment, 23*, 362–380. https://doi.org/10.1177/073428290502300405

Au, W., & Ferrare, J. J. (2015). Introduction: Neoliberalism, social networks, and the new governance of education. In W. Au & J. J. Ferrare (Eds.), *Mapping corporate education reform: Power and policy networks in the neoliberal state* (pp. 1-22). New York, NY: Routledge. https://doi.org/10.4324/9781315762401-1

August, D., Francis, D. J., Hsu, H. A., & Snow, C. E. (2006). Assessing reading comprehension in bilinguals. *The Elementary School Journal*, *107*, 221-238. https://doi.org/10.1086/510656

Ayers, J., & Owen, A. (2012). No Child Left Behind waivers: Promising ideas from second round applications. *Center for American Progress*. Retrieved from https://www.americanprogress.org/wp-ontent/uploads/issues/2012/07/pdf/nochildwaivers.pdf

Ball, S. J. (2018). Commercialising education: Profiting from reform! *Journal of Education Policy*, *33*, 587-589. https://doi.org/10.1080/02680939.2018.1467599

Ballou, D., & Springer, M. G. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher, 44*, 77-86. https://doi.org/10.3102/0013189X15574904

Bateman, B. (1991). Teaching word recognition to slow-learning children. *Journal of Reading, Writing, & Learning Disabilities International, 7*, 1-16. http://dx.doi.org/10.1080/0748763910070102

Beberman, M. (1962). The old mathematics in the new curriculum. *Educational Leadership, 19*, 373-375.

Begle, E. G. (1968). SMSG: The first decade. *The Mathematics Teacher, 61*, 239-245.

Berliner, D. (2005). The near impossibility of testing for teacher quality. *Journal of Teacher Education, 56*, 205-213. https://doi.org/10.1177/0022487105275904

Berliner, D. C. (2013). Problems with value-added evaluations of teachers? Let me count the ways! *The Teacher Educator*, *48*, 235-243. https://doi.org/10.1080/08878730.2013.827496

Berliner, D. C. (2014). Exogenous variables and value-added assessments: A fatal flaw. *Teachers College Record*, *116*(1), Article 17293. Retrieved from http://www.tcrecord.org/library

Berliner, D. C. (2017). The purported failure of America's schools, and ways to make them better [Equity Alliance blog]. Retrieved from http://www.niusileadscape.org/bl/the-purported-failure-of-americas-schools-and-ways-to-make-them-better-by-david-c-berliner/

Berliner, D. C., & Glass, G. V. (2014). *50 myths and lies that threaten America's public schools: The real crisis in education*. New York, NY: Teachers College Press.

Bond, L. (2005). *Throwing out the baby with the bath water*. Stanford, CA: Carnegie Foundation for the Advancement of Teaching. Retrieved from ERIC database. (ED498985)

Brady, M. P. (2019). An alternative, curriculum-based value-added model for teacher preparation programs: A research summary. *Educational Review*, 1-19. Published on-line first on July 17, 2019. https://doi.org/10.1080/00131911.2019.1601615

Brady, M. P., Duffy, M. L., Hazelkorn, M., & Bucholz, J. (2014). Policy and systems change: Planning for unintended consequences. *The Clearing House: A Journal of Educational Strategies, Issues, and Ideas, 87*(3), 102-109. https://doi.org/10.1080/00098655.2014.891882

Brady, M. P., Heiser, L.A., McCormick, J. K., & Forgan, J. (2016). Adding value-added models to teacher preparation programs: Validity and reliability threats, and a manageable alternative. *Educational Forum, 80*(3), 339-352. https://doi.org/10.1080/00131725.2016.1173150

Brady, M. P., & Miller, K. M. (2018). Curriculum-based VAM: An alternative to traditional VAM in clinically-rich teacher education. In D. Hoppey & D. Yendol-Hoppey (Eds.), *Outcomes of High-Quality Clinical Practice in Teacher Education* (pp. 105-126). Charlotte, NC: Information Age Publishing.

Brady, M. P., Miller, K., McCormick, J., & Heiser, L. A. (2018). A rational and manageable value-added model for teacher preparation programs. *Educational Policy, 32(5), 728-750*. https://doi.org/10.1177/0895904816673741. *(Advance online version published in 2016.)*

Brittingham, B. (2009). Accreditation in the United States: How did we get to where we are? *New Directions for Higher Education, 2009*(145), 7-27. https://doi.org/10.1002/he.331

Brownell, M. T., Rosenberg, M. S., Sindelar, P. T., & Smith, D.D. (2004). Teacher education: Toward a qualified teacher for every classroom. In A. M. Sorrells, H. J. Reith, & P. T. Sindelar (Eds.), *Critical issues in special education: Access, diversity, and accountability* (pp. 243-257). Boston: Allyn & Bacon.

Burke, L. M. (2016). Reducing federal intervention in education and moving toward student-centered policies: 10 steps for the incoming administration. *Backgrounder, 3177*, 1-6. Washington, DC: Heritage Foundation. Retrieved from http://report.heritage.org/bg3177

Buzick, H. M., & Laitusis, C. C. (2010). Using growth for accountability: Measurement challenges for students with disabilities and recommendations for research. *Educational Researcher, 39*, 537-544. https://doi.org/10.3102/0013189X10383560

Carnine, D., & Silbert, J. (1979). *Direct instruction reading*. Columbus, OH: Charles Merrill.

Chall, J. (1967). *The great debate*. New York, NY: McGraw Hill.

Chwalisz, K. (2003). Evidence-based practice: A framework for twenty-first-century scientist-practitioner training. *The Counseling Psychologist*, *31*, 497–528. https://doi.org/10.1177/0011000003256347

Cochran-Smith, M., Piazza, P., & Power, C. (2013). The politics of accountability: Assessing teacher education in the United States. *Educational Forum, 77*, 6-27. https://doi.org/10.1080/00131725.2013.739015

Collins, C. (2014). Houston, we have a problem: Teachers find no value in the SAS education value-added assessment system (EVAAS®). *Education Policy Analysis Archives, 22*(98). Retrieved from https://doi.org/10.14507/epaa.v22.1594

Cook, T. D. (2002). Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. *Educational Evaluation and Policy Analysis, 24*, 175-99. https://doi.org/10.3102/01623737024003175

Cook, B. G., & Cook, S. C. (2013). Unraveling evidence-based practices in special education. *The Journal of Special Education*, *47*, 71-82. https://doi.org/10.1177/0022466911420877

Cook, B. G., Buysse, V., Klinger, J., Landrum, T., J., McWilliam, R. A., Tankersley, M., & Test, D. W. (2015). CEC's standards for classifying the evidence base of practices in special education. *Remedial and Special Education, 36*, 220-234. https://doi.org/10.1177/0741932514557271.

Creemers, B. P. M., & Reezigt, G. J. (1996). School-level conditions affecting the effectiveness of instruction. *School Effectiveness and School Improvement, 7*, 197-228. https://doi.org/10.1080/0924345960070301

Cromwell, S. (2016). Whole language and phonics: Can they work together? *Education World.* Retrieved from https://www.educationworld.com/a_curr/curr029.shtml

Cuban, L. (1984). *How teachers taught: Constancy and change in American classrooms, 1890-1980* (1st ed.). [Research on Teaching Monograph Series]. Retrieved from ERIC database. (ED383498)

Cuban, L. (2004). The open classroom: Were schools without walls just another fad. *Education Next, 4*, 68-71.

Darling-Hammond, L. (2010). Teacher education and the American future. *Journal of Teacher Education*, *61*(1-2), 35-47. https://doi.org/10.1177/0022487109348024

Darling-Hammond, L., & Youngs, P. (2002). Defining "highly qualified" teachers: What does the "scientifically-based research" actually tell us? *Educational Researcher, 31*(9), 13-25. https://doi.org/10.3102/0013189X031009013

Demanding answers from the FL Department of Education. (n.d.). E.W. Scripps Company. Retrieved from https://www.omnivirt.com/view/24679

Deno, S. L., Fuchs, L. S. Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30*, 507-524.

Detrich, R., Keyworth, R., & States, J. (2016). Leveraging evidence-based practices: From policy to action. *Learning Disabilities: A Contemporary Journal*, *14*, 121-142.

Eaton J. S. (2015). *An overview of U.S. accreditation*. Washington, DC: Council for Higher Education Accreditation.

Engelmann, S., & Carnine, D. (1991). *Theory of instruction: Principles and applications*. Eugene, OR: ADI Press.

Every Student Succeeds Act of 2015*, 20 U.S.C. §§ 6301 et seq.* (2015).

Finn, C. E. (2013). Can digital learning transform education? First, we need a brand new K–12 system. *Education Next, 13*(1), 54-60. Retrieved from https://www.educationnext.org/files/ednext_20131_forum.pdf

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S.,…Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, *6*, 151–175. https://doi.org/10.1007/s11121-005-5553-y

Flesch, R. (1955). *Why Johnny can't read: And what you can do about it*. New York, NY: Harper & Row.

Floden, R. (2012). Teacher value added as a measure of program quality: Interpret with caution. *Journal of Teacher Education*, *63*, 356-360. https://doi.org/10.1177/0022487112454175

Florida Department of Education. (2018). PK-12 Public school data publications and reports. http://www.fldoe.org/accountability/data-sys/edu-info-accountability-services/pk-12-public-school-data-pubs-reports/staff.stml

Fuchs, D., & Deshler, D. D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research & Practice, 22*, 129–136. https://doi.org/10.1111/j.1540-5826.2007.00237.x

Fullan, M. (2011). *Choosing the wrong drivers for whole system reform.* Centre for Strategic Education, Seminar Series Paper No. 204. https://doi.org/10.4324/9781315762401-1

Garfolo, B. T., & L'Huillier, B. (2015). Demystifying assessment: The road to accreditation. *Journal of College Teaching & Learning, 12*, 151-169. https://doi.org/10.19030/tlc.v12i3

Giangreco, M. F., & Taylor, S. J. (2003). "Scientifically based research" and qualitative inquiry. *Research & Practice for Persons with Severe Disabilities*, *28*, 133-137. https://doi.org/10.2511/rpsd.28.3.133

Goldhaber, D. (2008). Teachers matter, but effective teacher policies are elusive. In H. Ladd & E. B. Fiske (Eds.), *Handbook of research in education finance and policy* (pp. 146-165). New York, NY: Routledge.

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher, 44*, 96–104. https://doi.org/10.3102/0013189X15575031

Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist, 6*, 126-135. https://doi.org/10.1080/19388076709556976

Greenberg, J., Putnam, H., & Walsh, K. (2014). *Training our future teachers: Classroom management* (Rev.). Retrieved from the National Council on Teacher Quality website: https://www.nctq.org/publications/home

Greenwood, C. R., & Maheady, L. (1997). Measurable change in student performance: Forgotten standard in teacher preparation? *Teacher Education and Special Education, 20*, 265–275. https://doi.org/10.1177/088840649702000307

Grossman, P., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal*, *45*(1), 184–205. https://doi.org/10.3102/0002831207312906

Head, B. (2016). Towards more 'evidence-informed' policymaking? *Public Administration Review, 76*, 472-484. https://doi.org/10.1111/puar.12475

Hempenstall, K. (n.d.). A history of disputes about reading instruction. Retrieved from the National Institute for Direct Instruction website: https://www.nifdi.org/news-latest-2/blog-hempenstall/396-a-history-of-disputes-about-reading-instruction

Henry, G. T., & Bastian, K. C. (2015). *Measuring up: The National Council on Teacher Quality's ratings of teacher preparation programs and measures of teacher performance.* Chapel Hill, NC: Education Policy Initiative at Carolina.

Hess, F. M. (2001). *Tear down this wall: The case for a radical overhaul of teacher certification.* Washington, DC: Progressive Policy Institute. Retrieved from http://www.aei.org/wp-content/uploads/2011/10/20070223_TearDownthisWallPPI.pdf

Holdheide, L., Goe, L., Croft, A., & Reschly, D. (2010). *Challenges in evaluating special education teachers and English Language Learner specialists.* Washington, DC: National Comprehensive Center for

Teacher Quality. Retrieved from http://www.gtlcenter.org/sites/default/files/docs/July2010Brief.pdf

Horntvedt, M. T., Nordsteien, A., Fermann, T., & Severinsson, E. (2018). Strategies for teaching evidence-based practice in nursing education: A thematic literature review. *BMC Medical Education*, *18*, 172-183. https://doi.org/10.1186/s12909-018-1278-z

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165–179. https://doi.org/10.1177/001440290507100203

Individuals with Disabilities Education Act (IDEA), 20 U.S.C. §§ 1400–1419 (2014).

Jones, N. D., Buzick, H. M., & Turkan, S. (2013). Including students with disabilities and English learners in measures of educator effectiveness. *Educational Researcher*, *42*, 234-241. https://doi.org/10.3102/0013189X12468211

Jorgensen, M. A., & Hoffmann, J. (2003). *History of the No Child Left Behind Act of 2001 (NCLB)* (Rev. 1). Upper Saddle River, NJ: Pearson Education. Retrieved from https://images.pearsonassessments.com/images/tmrs/tmrs_rg/HistoryofNCLB.pdf?WT.mc_id=TMRS_History_of_the_No_Child_Left_Behind

Keller, C., Brady, M. P., Duffy, M. L., Forgan, J., & Leach, D. (2008). If you build it and they still don't come: Playing the game of alternative certification. *Educational Forum, 72*, 228-244. https://doi.org/10.1080/00131720802046032

Kline, M. (1973). *Why Johnny can't add: The failure of the New Math.* New York, NY: St. Martin's Press.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, *34*(1), 26–38. https://doi.org/10.1177/0741932512452794

Kronholz, J. (2012). A new type of ed school: Linking candidate success to student success. *Education Next*, *12*(4), 2-8. Retrieved from https://www.educationnext.org/files/ednext_20124_kronholz.pdf

LaGrone, K., & Apthorp, M. S. (2017, August 7). *Hundreds of FL teachers fired over state test they keep failing: Unprecedented failures result from revised exam*. Scripps Media. Retrieved from https://www.fox4now.com/longform/hundreds-of-fl-teachers-fired-over-state-test-they-keep-failing

Lancaster, J., & Bain, A. (2019). Designing university courses to improve pre-service teachers' pedagogical content knowledge of evidence-based inclusive practice. *Australian Journal of Teacher Education, 44(*2), 51-65. http://dx.doi.org/10.14221/ajte.2018v44n2.4

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). Retrieved from http://www.tcrecord.org/Content.asp?ContentId=17294

Ledford, J. R., Barton, E. E., Hardy, J. K., Elam, K., Seabolt, J., Shanks, M., Hemmeter, M. L., & Kaiser, A. (2016). What equivocal data from single case comparison studies reveal about evidence-based practices in early childhood special education. *Journal of Early Intervention*, *38*, 79–91. https://doi.org/10. 1177/1053815116648000

Lewis, T. J., Hudson, S., Richter, M., & Johnson, N. (2004). Scientifically supported practices in emotional and behavioral disorders: A proposed approach and brief review of current practices. *Behavioral Disorders*, *29*, 247-259. https://doi.org/10.1177/019874290402900306

Lewis, W. D., & Young, T. V. (2013). The politics of accountability: Teacher education policy. *Educational Policy*, *27*, 190-216. https://doi.org/10.1177/0895904812472725

Lincove, J. A., Osborne, C., Dillon, A., & Mills, N. (2014). The politics and statistics of value-added modeling for accountability of teacher preparation programs. *Journal of Teacher Education, 65*(1), 24–38. https://doi.org/10.1177/0022487113504108

Lingard, B. (2013). The impact of research on education policy in an era of evidence-based policy. *Critical Studies in Education*, 54, 113–131. https://doi.org/10.1080/17508487.2013.781515

Maddox, K., & Feng, J. (2013, October). *Whole language instruction vs. phonics instruction: Effect of reading fluency and spelling accuracy of first grade students*. Paper presented at the Georgia Educational Research Annual Conference, Savannah, GA.

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy, 4*, 572-606. https://doi.org/10.1162/edfp.2009.4.4.572

McCall, R. B., & Green, B. L. (2004). Beyond the methodological gold standards of behavioral research: Considerations for practice and policy. *Society for Research in Child Development Social Policy Reports*, *18*, 1-19. https://doi.org/10.4324/9781315762401-1

McCormick, J. K., Brady, M. P., Miller, K. M., Heiser, L. H., & Morris, J. D. (2018). *A curriculum-based alternative value-added model for special education teacher preparation programs*. Manuscript submitted for publication.

McCormick, J., Brady, M. B., Morris, J. D., Heiser, L. A., & Miller, K. (2019). Further examination of a curriculum-based Value-Added Model for teacher preparation: Exploring the role of teachers' behavior on K-12 student learning. *The Teacher Educator, 54*(1), 60-71. https://doi.org/10.1080/08878730.2018.1512023

McGuire, P. A. (2009). Accreditation's benefits for individuals and institutions. *New Directions for Higher Education, 2009*(145), 29-36. https://doi.org/10.1002/he.332

McKnight, L., & Morgan, A. (2019). A broken paradigm? What education needs to learn from evidence-based medicine. *Journal of Education Policy*. https://doi.org/10.1080/02680939.2019.1578902

Miller, J. W. (1990). What happened to New Math? *American Heritage, 40*(8), 76-83.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Department of Education. Retrieved from ERIC database. (ED226006)

National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Retrieved from the U.S. Department of Health and Human Services website: https://www.nichd.nih.gov/publications/pubs/nrp/smallbook

No Child Left Behind Act of 2001, *Pub. L. No. 107-110* 115 Stat. 1425 (2001).

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children*, *71*, 137-148. https://doi.org/10.1177/001440290507100201

Papay, J. P. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Education Research Journal, 48*, 163-193. https://doi.org/10.3102/0002831210362589

Perrone, V. (1972). *Open education: Promise and problems*. Phi Delta Kappa Educational Foundation, Fastback Series, No. 3. Retrieved from ERIC database. (ED062726)

Phi Delta Kappan. (2017). The 49th annual PDK poll of the public's attitudes toward the public schools: Academic achievement isn't the only mission. *Phi Delta Kappan*, *99*(1), NP1–NP32. https://doi.org/10.1177/0031721717728274

Ravitch, D. (2013). *Reign of error: The hoax of the privatization movement and the danger to America's public schools*. New York, NY: Alfred A. Knopf.

Rothenberg, J. (1989). The open classroom reconsidered. *Elementary School Journal, 90*, 69-86. https://doi.org/10.1086/461603

Rowe, E. E., & Skourdoumbis, A. (2019). Calling for 'urgent national action to improve the quality of initial teacher education': The reification of evidence and accountability in reform agendas. *Journal of Education Policy*, *34*(1), 44-60. https://doi.org/10.1080/02680939.2017.1410577

Rueter, J. A., & Simpson, C. G. (2012). The promises and realities of evidence-based practices: Perceptions from assessment personnel. *Administrative Issues Journal: Education, Practice, and Research*, *2*, 122-132. https://doi.org/10.5929/2011.2.1.3

Rycroft-Malone, J., Seers, K., Titchen, A., Harvey, G., Kitson, A., & McCormack, B. (2004). What counts as evidence in evidence-based practice? *Journal of Advanced Nursing*, *47*(1), 81–90. https://doi.org/10.1111/j.13652648.2004.03068.x

Scheeler, M. C., Budin, S., & Markelz, A. (2016). The role of teacher preparation in promoting evidence-based practice in schools. *Learning Disabilities: A Contemporary Journal*, *14*, 171-187.

Schneider, J. (2018). Marching forward, marching in circles: A history of problems and dilemmas in teacher preparation. *Journal of Teacher Education*, *69*, 330-340. https://doi.org/10.1177/0022487117742904

Shores, R. E. (1979). Evaluation and research. *Teacher Education and Special Education, 2*(3), 68–71. https://doi.org/10.1177/088840647900200326

Silberman, C. E. (1971). *Crisis in the classroom: The remaking of American Education*. New York, NY: Random House.

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children*, *31*, 351-380. https://doi.org/10.1353/etc.0.0007

Smith, F. (2004). *Understanding reading: A psycholinguistic analysis of reading and learning to read* (6th ed.). Mahwah, NJ: Lawrence Erlbaum.

Stahl, S. A., & Kuhn, M. R. (1995). Does whole language or instruction matched to language styles help children to read? *School Psychology Review, 24*, 393-405.

Steinbrecher, T. D., Selig, J. P., Cosbey, J., & Thorstensen, B. I. (2014). Evaluating special educator effectiveness: Addressing issues inherent to Value-Added Modeling. *Exceptional Children, 80*, 323-336. https://doi.org/10.1177/0014402914522425

Strain, P. S. (2018). Personal thoughts on early childhood special education research: An historical perspective, threats to relevance, and call to action. *Journal of Early Intervention*, *40*, 107-116. https://doi.org/10. 1177/1053815117750411

Swank, P. R., Taylor, R. D., Brady, M. P., & Freiberg, J. (1989). Sensitivity of classroom observation systems: Measuring teacher effectiveness. *Journal of Experimental Education, 57*, 171-186. https://doi.org/10.1080/00220973.1989.10806504

Vigdor, J. (2013). Solving America's math problem. *Education Next*, *13*(1). Retrieved from https://www.educationnext.org/solving-america%E2%80%99s-math-problem/

Walsh, K., & Hale, C. (2004). *Increasing the odds: How good policies can yield better teachers.* Washington, DC: National Council on Teacher Quality. Retrieved from https://www.nctq.org/nctq/images/nctq_io.pdf

Whitehurst, G. J. (2012). The value of experiments in education. *Education Finance and Policy, 7*, 107-123. https://doi.org/10.1162/EDFP_a_00058

Yell, M. L., Conroy, T., Katsiyannis, A., & Conroy, T. (2013). Individualized education programs (IEPs) and special education programming for students with disabilities in urban schools. *Fordham Urban Law Journal*, *41*, 669-714.

About the Authors

Michael P. Brady
Florida Atlantic University
mbrady@fau.edu
http://orcid.org/0000-0002-3240-796X
Michael Brady, Ph.D., is a Professor and Chair of the Department of Exceptional Student Education at Florida Atlantic University, with administrative responsibility for programs in early childhood, special education, and the *Academy for Community Inclusion*, FAU's post-secondary program for adults with developmental disabilities. His scholarship incudes teacher and faculty preparation, effective teaching interventions, policy and reform initiatives, social relationships between people with and without disabilities, and community inclusion of people with intellectual disabilities.

Michael Hazelkorn
K & M Coastal Consultants
mhazelkorn@ccga.edu
Michael Hazelkorn has a master's degree in special education from the University of Arizona and a doctorate in special education from the University of Georgia. He taught at the University of Wisconsin-Eau Claire, the University of West Georgia, and the College of Coastal Georgia. He served as chair of a department at two universities and was the dean of the school of education at the third. He is now retired.

education policy analysis archives

Volume 27 Number 137 October 28, 2019 ISSN 1068-2341

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

Join EPAA's Facebook community at https://www.facebook.com/EPAAAAPE and **Twitter feed** @epaa_aape.

archivos analíticos de políticas educativas
consejo editorial