

Kimberly Kappler Hewitt • Audrey Amrein-Beardsley
Editors

Student Growth Measures in Policy and Practice

Intended and Unintended Consequences
of High-Stakes Teacher Evaluations
**When Theoretical Models Meet School Realities: Educator Responses to Student
Growth Measures in an Incentive Pay Program**

palgrave
macmillan

Editors

Kimberly Kappler Hewitt
University of North Carolina
at Greensboro
Greensboro, USA

Audrey Amrein-Beardsley
Arizona State University
Tempe, Arizona, USA

ISBN 978-1-137-53900-7

ISBN 978-1-137-53901-4 (eBook)

DOI 10.1057/978-1-137-53901-4

Library of Congress Control Number: 2016950596

© The Editor(s) (if applicable) and The Author(s) 2016

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use. The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Cover illustration: © J Stromme / Alamy Stock Photo

Printed on acid-free paper

This Palgrave Macmillan imprint is published by Springer Nature
The registered company is Nature America Inc. New York

“First, Do No Harm?”: A Framework for Ethical Decision-Making in Teacher Evaluation

*Carla M. Evans, Jade Caines Lee,
and Winston C. Thompson*

In the past five years, seismic shifts have taken place within teacher evaluation policies (Collins & Amrein-Beardsley, 2014; Doherty & Jacobs, 2015). Incentivized by the federal government through Race-to-the-Top funds and No Child Left Behind (NCLB) waivers, student growth measures (SGMs) are now widely used as critical components in determining teacher effectiveness throughout the United States, despite concerns regarding their psychometric properties and how they are applied within high-stakes educational decision-making (Amrein-Beardsley, 2014; Baker et al., 2010; Berliner, 2014; Koedel & Betts, 2009; Newton, Darling-Hammond, Haertel, & Thomas, 2010; Reardon & Raudenbush, 2009; Scherrer, 2011). While studies have suggested that SGMs may not accurately capture a teacher’s effect on student achievement outcomes, decisions related to teacher tenure, compensation, dismissal, and promotion are increasingly being based on SGMs

C.M. Evans (✉) • J.C. Lee • W.C. Thompson
University of New Hampshire, Durham, NH, USA

(Amrein-Beardsley & Collins, 2012; Baker, Oluwole, & Green, 2013; Lavigne, 2014; Paige, 2012, 2014; Pullin, 2013).

Alongside the methodological and social critiques of SGMs found in the other chapters of this book, there are ethical concerns that also, to date, may not be adequately considered by educational decision-makers. Hence, it may be tempting to engage these ethical concerns by borrowing the ubiquitous phrase, "First, do no harm," from the medical field. However, the decision-making landscape in both medicine and education may be more complicated than that adage would suggest. Making decisions about the design and/or implementation of teacher evaluation systems without considering the intended and unintended consequences, especially ethical ones, is ill-advised. No matter which methodologies are used to assess teacher quality, some account of moral priorities/values is also implicitly or explicitly endorsed by mere adoption (House & Howe, 1999; House, 1978, 1980; Scriven, 1967; Strike, 1980; Stufflebeam, 1994). Therefore, the complexity of ethical issues must be considered alongside psychometric frameworks when assessing teacher quality (Kane, 2006, 2013; Messick, 1980, 1985, 1989; Newton & Shaw, 2014). Decision-makers would do well to consider this overlooked moral dimension in their deliberations about teacher evaluation reforms (Elwood, 2013; Evans, 2015).

Acknowledging the need to explore the moral dimensions of teacher evaluation, the purpose of this chapter is to provide a practical framework that decision-makers can use to contextualize, analyze, and more thoughtfully navigate ethical quandaries. More specifically, we argue that when designing and/or implementing teacher evaluation systems, decision-makers must weigh all the relevant factors, including ethical ones. Our framework assists decision-makers in this process by providing critical steps for a comprehensive and systematic consideration of relevant ethical issues.

The chapter is organized into two sections. In the first section, in order to set the context, we define ethical dilemmas. This helps to preview the ethical terrain related to the design and/or implementation of teacher evaluation systems. Next, we detail a hypothetical year-in-the-life of an average educational decision-maker who is faced with numerous ethical dilemmas related to the use of SGMs in teacher evaluation systems. This hypothetical example, referred to as *The Westview Dilemmas*, provides an ethical perspective on evaluation and educational decision-making. In the second section, we provide a useful framework that might guide (not dictate) educational decision-makers and their judgments. We utilize *The*

Westview Dilemmas to demonstrate how the framework can be applied when evaluating the ethical design and/or implementation of teacher evaluation systems.

SECTION I: ETHICAL DILEMMAS

An ethical dilemma is a clash between the right and the right; it is the tension created when two ethical principles conflict. In other words, an ethical dilemma is one in which it is truly unclear what ought to be done. In the context of schooling, educational decision-makers may face ethical dilemmas more often than dichotomous scenarios of right and wrong. The reality of decision-making is complicated, especially when it involves potential high-stakes consequences for teachers. For example, as an educational decision-maker in today's high-stakes accountability climate, it can be difficult to determine the "right" course of action when evaluating teacher effectiveness. This is why weighing ethical impacts is critical when designing and/or implementing teacher evaluation systems. In order to explore some of the ethical tensions that occur when SGMs are included as a component in teacher evaluation systems, we present a hypothetical year-in-the-life of an average educational decision-maker. As you read, imagine yourself in this situation: How and on what basis would you respond to the issues raised?

The Westview Dilemmas

Although you have only recently been hired as superintendent of the Westview School District, you are quickly beginning to realize that the newly redesigned teacher evaluation system, which now includes SGMs, is a major issue for teachers in your district.

One Morning in September

You've spent most of your morning answering a deluge of emails, calming concerns from apprehensive teachers. Some individual teachers are still frustrated because the findings from last year's pilot evaluations seem unclear. They are unsure how they are supposed to use their numerical scores or classification ratings to improve their teaching. Some wonder whether the evaluation results will impact their salaries and even job security during this, the first year of high-stakes implementation. Many are quite sure that these "simplistic" scores and effectiveness ratings do not

accurately capture all that they do for their students. Though the academic year has just started, these teachers are insisting on conclusive answers. You sigh heavily as you begin to type yet another response.

Two Months Later

In addition to the steady stream of emails that you are still getting from individual teachers, you are also receiving petitions from groups of teachers within the district. For instance, English language, gifted, and special education teachers are convinced that measuring their effectiveness through SGMs is unfair and doesn't adequately represent their quality as teachers. They wish to opt out of having SGMs on their evaluations and are gathering signatures to demonstrate support for their proposed changes. You place these requests in a growing folder on your computer.

The Month of Testing

Today's lunch meeting was the third this month in which a principal mentions currents of dissatisfaction from groups of teachers in district schools. According to this principal, some teachers feel that the strong focus on standardized test preparation detracts from their ability to do much else throughout the school year. These teachers are worried that their activities might not align with their vision of what "good teaching is all about." They simultaneously worry about the professional consequences of deviating from these preparations. Their principal is unsure of what to tell them. Hearing this, you frown and slowly shake your head as you leave for your next meeting.

The Last Day of School Before Summer

This afternoon's visit to a few schools in the district has not gone quite as planned. While it seemed like a great idea to get a "view from the ground" on the last school day, you had not considered the fact that teacher evaluation results would be distributed today. Many teachers, including some well-known veterans, are visibly upset. Some have tears in their eyes, and one makes an especially strong statement about being "fed up with the whole thing!"

You are happy to return to your quiet office now, as you really need a moment to think through all that you experienced today. As you sit in your chair, your assistant knocks on your door to tell you that a reporter from the *Westview Gazette* is on the phone. Apparently, the newspaper would like you to comment on the teacher evaluation results. More specifically, the reporter wants to know how you are going to protect students

from “bad” teachers. You close your eyes as you place your hand on the telephone receiver on your desk.

SECTION II: A FRAMEWORK FOR ETHICAL DECISION-MAKING

While the snapshots included in *The Westview Dilemmas* are not meant to be exhaustive or representative of every superintendent’s experience, they are examples of some of the challenges educational administrators may face in designing and/or implementing teacher evaluation systems in which SGMs are a key component. The superintendent faces one ethical challenge after another; meanwhile, tensions build without an organized approach to handling these conflicts. Since the superintendent has flexibility in how the Westview School District incorporates SGMs into the teacher evaluation system, a framework for navigating the complicated ethical terrain may be useful. Given these complexities, as well as external accountability pressures, we argue for a practical framework that will aid decision-makers in dealing with these types of ethical tensions. This framework can be helpful in navigating a complicated terrain, while illuminating possible considerations for high-stakes educational decision-making.

In our *Practical Framework and Feedback Mechanism for Ethical Decision-Making in Teacher Evaluation* (Fig. 9.1), we lay out a process for thinking through and evaluating the ethical impacts of teacher evaluation, especially in decision-making relative to SGMs. This framework includes the following: (1) *identifying* the ethical issues at stake in the use of SGMs using ordinary moral intuitions, (2) *analyzing* those ethical issues using the best available relevant research and professional standards/codes, (3) *applying* five ethical principles to illuminate ethical ramifications, and (4) *reflecting* on the process and re-examining (if necessary) final decision-making.

Utilizing *The Westview Dilemmas*, we turn back now to the framework to examine the ethical issues that arise in the implementation of Westview School District’s teacher evaluation system, in this hypothetical (but likely) case and context.

Step 1: Identify Ethical Issues

The task of identifying ethical conflicts includes the application of ordinary moral intuitions (Kitchener, 1984). This process is akin to cataloging the potential conflicts and concerns relative to the design and/

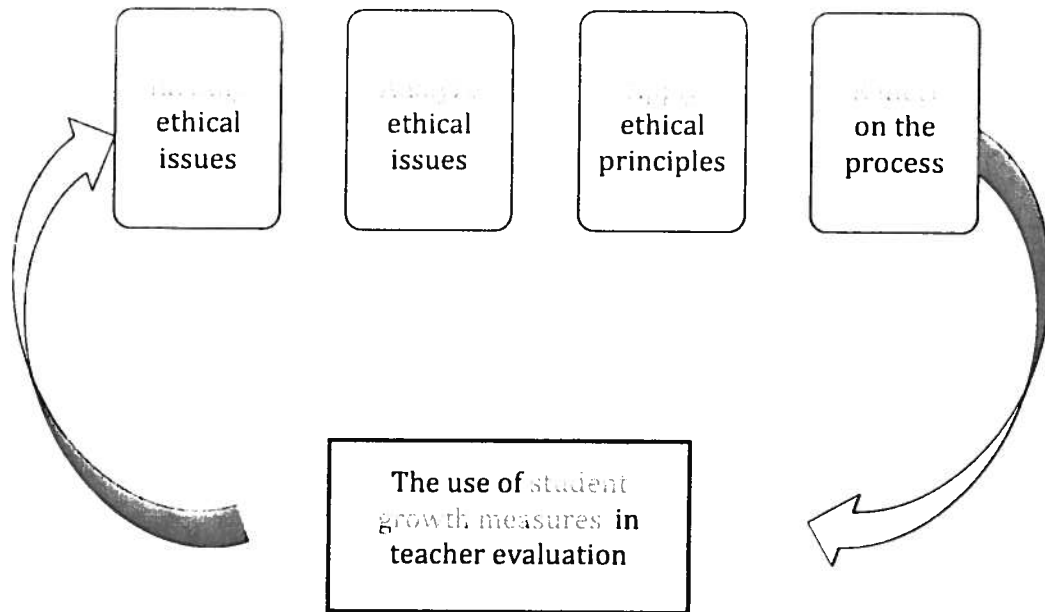


Fig. 9.1 A practical framework and feedback mechanism for ethical decision-making in teacher evaluation

or implementation of SGMs in teacher evaluation. At its core, this step asks: “What conflicts and concerns exist, or could potentially exist, from this course of action or decision?” Examining *The Westview Dilemmas*, the basic conflict involves the collision between the promises of teacher evaluation policy changes (e.g., a more effective teacher workforce and increased student achievement outcomes) with the impact of such policies on those at their receiving ends. As such, a few ethical issues are immediately apparent. Teachers may be distrustful of the results of the evaluation based on SGMs, and many may be concerned about negative impacts of evaluation results. Teachers may also be concerned about their loss of professional judgment and feel demoralized by evaluation results. Potential breaches of teacher confidentiality and privacy leave the superintendent in a difficult situation.

While this step may seem basic and perfunctory, it is important that educational decision-makers catalog the ethical issues that may occur in designing and/or implementing a teacher evaluation system using SGMs. If the conflicts and concerns are unknown, they cannot be proactively addressed, preferably during the planning phase before high-stakes implementation begins.

Step 2: Analyze Ethical Issues

Having first identified ethical issues, a good decision-making process requires a thorough analysis of those issues. In order to do this, educational decision-makers could examine (1) relevant research and (2) professional standards/codes. First, when examining relevant research, it would be ideal to conduct a focused analysis of the research literature. This could include reading published literature on the intended and unintended consequences linked to SGMs in teacher evaluation. However, it may be most useful for decision-makers to locate a few different sources that synthesize the literature on impacts, so that they may have a critical overview of the research writ large (e.g., Amrein-Beardsley, 2014; Harris, 2011; Lavigne & Good, 2013).

Next, in analyzing ethical issues, a decision-maker could examine professional standards and/or codes. Given the varied types of educational decision-making organizations that currently exist, many professional standards or codes may guide the design and/or implementation of teacher evaluation systems. For example, the *Personnel Evaluation Standards* (Gullickson & Joint Committee on Standards for Educational Evaluation, 2009) and the American Evaluation Association's *Guiding Principles for Evaluators* (American Evaluation Association, 2004) may be two professional codes from the evaluation field that have direct relevance for teacher evaluation. Additionally, a *Bill of Rights for Teacher Evaluation* (Strike, 1990; Strike & Bull, 1981) works in concert with both sets of guidelines to set standards, principles, and rights for the process of teacher evaluation. And yet, these professional codes are limited in that they provide no direct guidance as to what ought to be done when standards, principles, or rights conflict, or are ambiguous (Morris, 2008; Newman & Brown, 1996; Simons, 2006). Therefore, despite the *necessity* of professional standards/codes, they cannot be wholly *sufficient* for ethical decision-making relative to the use of SGMs in teacher evaluation systems (Newman & Brown, 1996).

Step 3: Apply Ethical Principles

Once ethical issues are identified and analyzed, the application of ethical principles can be useful when judging the merit, worth, or quality of teacher evaluation measures. They do this by (1) sharpening ordinary moral sensibilities, (2) giving comprehensive and reliable insight into the

full scope of moral issues that must be considered, and (3) serving as a useful rationale for ethical decision-making (Kitchener, 1984; Morris, 2003, 2011). While there are several ways to conceptualize the use of ethics in educational evaluation (Bunda, 1985; House, 1976; Morris, 2008; Newman & Brown, 1996; Simons, 2006; Strike, 1979, 1990), we utilize the received model of applied ethics from the social sciences, which can help explain how individuals should be treated and on what basis decisions should be evaluated (Beauchamp & Childress, 2001; Bloch & Green, 2006; Drane, 1982; Howe & Moses, 1999; Kitchener & Kitchener, 2009, 2012). We draw on several general-usage definitions of the term ethics: a) rules of conduct developed by, and for, members of a particular profession, b) principles of morality, especially those detailing what is right or wrong with an action, and c) the science of the study of ideal human behavior (Newman & Brown, 1996). We treat ethics as a disciplinary perspective in order to illuminate ethical issues within current teacher evaluation reform. More specifically, we deal with several ethical principles that we believe to be especially relevant for teacher evaluation: nonmaleficence, beneficence, autonomy, justice, and fidelity (Table 9.1).

These principles offer a flexible approach to moral justification in teacher evaluation, while providing a good source of arbitration in moments of moral dilemma. In applying these ethical principles to *The Westview Dilemmas*, we do not view each principle as unrelated. Also, our purpose is not to give an exhaustive treatment of examples relative to how the principles and SGMs may intersect. Instead, we aim to provide adequate detail for an understanding of the ethical principles and demonstrate how educational decision-makers might apply this understanding to the design and/or implementation of teacher evaluation systems. In what follows, we explore some of the ethical issues and tensions in *The Westview Dilemmas* (see Table 9.2).

Nonmaleficence. Nonmaleficence is the avoidance of doing harm to others. In other words, this principle suggests that certain kinds of teacher evaluation activities should be forbidden if they unduly harm teachers. *Harm* is defined as inflicting intentional pains or risking pains to others. However, what constitutes *undue* harm or risk is not always easily discernable (Newman & Brown, 1996). Some traditions of ethical theory argue that “First, do no harm” should be considered the ultimate criterion for ethical decision-making, while rival traditions have suggested that the risk of harming others may, at times, be justifiable.

It is clear from *The Westview Dilemmas* that teachers have been harmed; they are stressed, demoralized, and fear for their futures (see Table 9.2).

Table 9.1 Key characteristics of ethical principles related to teacher evaluation

<i>Ethical principle</i>	<i>Key characteristics</i>	<i>Questions</i>
	<i>Teacher evaluation policies should...</i>	
Nonmaleficence	...not cause harm to teachers. That is, certain kinds of teacher evaluation activities should be forbidden if they unduly harm teachers.	What undue harm to teachers (psychological or actual) is likely to occur as a result of the decision and action?
Beneficence	...do good and benefit teachers and key stakeholders, as well as balance potentially beneficial consequences against the potentially harmful ones.	What good can come to teachers and key stakeholders as a result of the decision and action? What are the potential benefits and risks of action as compared to the potential benefits and risks of inaction?
Autonomy	...recognize that teachers deserve respect and are possessed of rights.	Are any teachers' rights affected?
Justice	...be fair. There should be established rules and procedures for determining teacher effectiveness that are both accurate and conducted in a fair manner.	What issues are related to fairness and accuracy in teacher evaluation? Are multiple perspectives being gathered, or multiple measures used?
Fidelity	...be honest and trustworthy. Teacher confidentiality and privacy should also be honored with regards to evaluation results.	What contractual obligations apply, and are they being fulfilled?

Note: Based on the work of Kitchener and Kitchener (2012) and Newman and Brown (1996)

However, the critical issue with regards to nonmaleficence is whether they have been *unduly* harmed. Just because a teacher is upset about receiving a negative evaluation does not mean she/he was unethically treated. But, if it is the case that some teachers were misclassified or erroneously evaluated, then those teachers *have* been unduly harmed and actions may need to be revisited.

Beneficence. Beneficence suggests there are certain positive obligations to do good or benefit others (Kitchener, 1984). As such, it contrasts with the principle of nonmaleficence, as that principle *forbids* certain kinds of actions and decisions if they unduly harm teachers. Consequently, beneficence is often conceptualized in two ways: (1) helping others by promoting their good, and (2) protecting others from harm by balancing potentially beneficial consequences against potentially harmful ones.

In *The Westview Dilemmas*, a reporter from the local newspaper wants to know what the superintendent will do to protect students from sub-

Table 9.2 Potential ethical issues in the Westview dilemmas organized by time stamp

<i>Potential ethical issues</i>	<i>Addressed issues</i>	<i>Other issues not addressed</i>
<p><i>One morning in September</i> Findings from evaluations are unclear Unsure how to use evaluation to improve teaching Teachers wonder about the impact of results on their salaries and job security Evaluations do not accurately capture all that they do for their students</p>	<p><i>Nonmaleficence:</i> Risk of misclassification</p>	<p><i>Justice:</i> Lack of transparency and formative use</p> <p><i>Justice:</i> Relevant evaluation criteria <i>Fidelity:</i> Accurate reflection of performance</p>
<p><i>Two months later</i> Groups of teachers are convinced SGMs are unfair and don't adequately represent their quality as teachers</p>	<p><i>Justice:</i> Lack of fair and equitable application</p>	<p><i>Nonmaleficence:</i> Risk of misclassification <i>Justice:</i> Relevant evaluation criteria <i>Fidelity:</i> Accurate reflection of performance</p>
<p><i>The month of testing</i> Teachers are dissatisfied with the testing climate and its impact on their vision of what "good teaching is all about" Teachers are concerned about narrowing of the curriculum, as well as the consequences of not doing enough test preparation</p>	<p><i>Autonomy:</i> Professional judgment</p>	<p><i>Nonmaleficence:</i> Stress and demoralization</p> <p><i>Beneficence:</i> Helping others vs. protecting others from harm</p>
<p><i>The last day of school before summer</i> Teachers are visibly upset on receiving evaluation reports Newspaper wanting comment on the teacher evaluation results, especially protecting students from "bad" teachers</p>	<p><i>Nonmaleficence:</i> Stress and demoralization <i>Fidelity:</i> Privacy and confidentiality <i>Beneficence:</i> Helping others vs. protecting others from harm</p>	

par teachers (see Table 9.2). Oftentimes, in discussions about teacher evaluation, the potential negative consequences for teachers are pitted against the potential negative consequences for students taught by ineffective teachers (cf., Goldhaber, 2010, pp. 24–25). In other words, should

teachers shoulder the increased risk of harm in evaluations, or should that risk be borne by students through inadequate instruction? This dichotomy, however, is false. It is not *who* should shoulder the risk in evaluation, but *how* might the evaluation be designed to accurately reflect the full range of influences on student achievement while recognizing the limited impact that teachers may have on increasing students' standardized test scores (American Statistical Association, 2014; Amrein-Beardsley, 2014; Berliner, 2013; Schochet & Chiang, 2012). Fulfilling the ethical principle of beneficence entails attempting to act in accord with our positive obligations while protecting others from harm.

Autonomy. Autonomy, which entails equal respect for persons, recognizes that teachers deserve respect and are possessed of rights. This principle typically focuses upon two related aspects: freedom of action and freedom of choice (Kitchener, 1984). In other words, people should be free to act and choose in accordance with their own free will except in situations in which their autonomous choice interferes with someone else's similar freedom to act and choose. The principle of autonomy assumes that teachers are competent enough to make decisions about how best to teach students and that the evaluation process ought not needlessly impede teachers' freedom to teach students according to their best professional judgment.

In *The Westview Dilemmas*, however, the use of SGMs in teacher evaluations creates a restrictive environment for some teachers. For example, some teachers express concerns to their principal about narrowing of the curriculum and teaching in a manner contrary to the best methods indicated by their professional judgment (see Table 9.2). The teachers may not be basing their curricular decisions on student learning needs but, rather, may be acting from the fear that if they do not focus considerable time and energy on test preparation, they may lose their jobs. This example highlights how a teacher's autonomy may be compromised in an evaluation system that links teachers' compensation, promotion, dismissal, or tenure decisions to student achievement test scores.

Justice. Justice in this context is operationalized as procedural fairness. This fairness requires that there be established rules and procedures for determining teacher quality and that these be reasonable, rather than arbitrary or capricious (Strike & Bull, 1981). In the context of teacher evaluation, justice necessitates the consideration of the rights and concerns of teachers. In general, ensuring the procedural fairness of an evaluation often involves collecting multiple stakeholder perspectives and utilizing multiple measures in order to validate results.

As evidenced in *The Westview Dilemmas*, the nonrandom sorting of students into classrooms has the potential to make teacher evaluation systems that use SGMs unfair for certain groups of teachers (see Table 9.2). For example, teachers who have inordinate numbers of homogenous groupings of students, such as English language learners and gifted and special needs students, are differentially impacted by evaluations based on student achievement tests scores because of the students they teach (Amrein-Beardsley & Collins, 2012; Berliner, 2014; Hill, Kapitula, & Umland, 2011; Newton et al., 2010; Rothstein, 2009). The principle of justice highlights ethical impacts related to the use of SGMs in teacher evaluations because it reminds us that teachers must be treated equitably and fairly. It is morally indefensible to knowingly evaluate groups of teachers using criteria that are unfair. The ethical principle of justice cautions against evaluating and comparing teachers based on the students they teach.

Fidelity. Fidelity, originating from the Latin word, *fidelitas*, means faithful. As such, fidelity connotes the honesty and trustworthiness of evaluation processes and results. From considering the ethical principle of fidelity, a decision-maker might engage issues of privacy and publicity. For example, should evaluation results be shared with anyone beyond the individual teacher? Is it defensible to share evaluation results among a group of teachers such as grade-level peers, or at the school- or district-level? Is there a reason why evaluation results should be shared with parents and/or members of the community?

While there is certainly a public desire for information about teacher effectiveness, publishing evaluation ratings on individual teachers in the *Westview Gazette* may exemplify a fidelity breach (see Table 9.2). One may argue that evaluation information needs to be public because parents and members of the community have a right to know if competent teachers are teaching their students. On the other hand, one may argue from the principle of fidelity that teachers should be able to trust that their privacy and confidentiality are not compromised in the quest for public accountability.

Tensions between the principles. In summary, applying these five ethical principles provides a foundation for a comprehensive and reliable framework to judge the use of SGMs in teacher evaluation. Each principle sharpens ordinary moral intuitions and provides insight into the full scope of moral issues that must be considered in making better decisions when in ethical doubt. This does not mean, however, that the ethical principles will not conflict, especially in complicated situations. Ethical dilemmas are, after all, clashes between the right and the right. In *The Westview*

Dilemmas, for example, many different principles are in tension with one another. From one perspective, not publishing names of "bad" teachers (fidelity) is in tension with holding the school district accountable for weeding out incompetent teachers in order to do "good" for students (beneficence). From another perspective, evaluating all teachers equitably and fairly (justice) is in tension with a desire to link teacher evaluations to student achievement test scores (beneficence).

Ethicists have conceptualized how to handle the tension between principles in different ways. For some, *balancing* mutually conflicting principles is seen as the "ultimate ethical act" (House, 1993, p. 168; c.f., Dworkin, 1979; Kymlicka, 1990; Rawls, 1971). For others, particular ethical principles should be *traded-off* or *prioritized* above others. Ross (1930), for example, prioritizes nonmaleficence above the other principles by asserting that, *ceteris paribus*, the avoidance of doing harm to another, is superior to doing good for others. Nozick (1974), on the other hand, would accord regulative primacy to the principle of autonomy on the grounds that individuals have inviolable autonomy rights that cannot be traded for societal benefits.

As we noted in the section on ethical dilemmas, however, ethical dilemmas are inherently complex. This implies that searching for the *one right way* to contain the tension between conflicting ethical principles and for the *one right answer* to an ethical dilemma in evaluation is ill-advised. As Morris (2008) says, "There may be a number of right answers, each one representing a different combination of ethical pros and cons" (Morris, 2008, p. 14). We believe that in highlighting the tensions between mutually conflicting ethical principles, one can better understand the ethical ramifications of a certain course of action and thereby make better-informed, ethically sensitive decisions. At minimum, we hope such ethical principles might aid in asking the following critical questions: What ethical price must we pay in order to achieve the policy objective in this situation, and is that price too steep? (Morris, 2003).

Step 4: Reflect on the Process

Thus far, our framework for ethical decision-making has suggested practical action steps for educational decision-makers engaged with ethical dilemmas. While the identification, analysis, and application of ethical considerations may seem a straightforward task in principle, it is work that is rife with complexities and ambiguities in practice. Educational

decision-makers must reflect on these difficulties if they are to make better decisions in the present and the future.

At this stage in the framework, educational decision-makers must resist two missteps: (1) the full certainty that they have reached a position from which they can act well, and (2) the conviction that they now possess a static skill for doing so in the future. In reflecting on their abilities to make better decisions about the use of SGMs in the design and/or implementation of teacher evaluation systems, educational decision-makers would do well to resist the hubris of certainty. Even as one moves closer to the point of a decision, there is significant value in humbly returning to previous steps of the framework. To do so is to accept the possibility that perspectives gained through the later steps may reveal new dimensions to previous considerations. A reflection on that process can serve as a fail-safe against earlier biases, misunderstandings, and mistakes, to say nothing of improving perceptions of present challenges. This reflection also aids in reinforcing a similar measure of humility in relation to future dilemmas. One ought not assume that, having avoided ethical pitfalls in one decision, a quickly reached decision is guaranteed in the future. Given the complexities of these ethical issues, as well as the high-stakes contexts in which these issues currently exist, the steps of the framework must be re-engaged, and the educational decision-maker must treat each dilemma as unique. In reflecting on the process, either to avoid a regrettable decision in the present or a reflexive response in the future, the educational decision-maker must consider at least these questions:

- What issues am I not attending to? Why?
- Have I considered all the relevant issues?
- In what ways does new information allow me to better see the scope of the dilemma?
- Am I inclined to certain decisions, and do those inclinations prioritize some principle or group?
- Am I acting in good faith through the steps of framework?
- In making this decision, what am I trying to accomplish?

These are some questions with which educational decision-makers can engage in order to reflect on their process, but they are just a start. Reflection questions may expand as an educational decision-maker becomes more comfortable in this type of inquiry. Moreover, the reflection process naturally initiates a feedback mechanism on the use of SGMs

in teacher evaluation. As reflection ensues, those reflections are contrasted with alternative courses of actions, including maintaining the *status quo*. Overall, this step keeps educational decision-makers sensitive to the ethical dimensions of this type of evaluation.

In considering a framework for ethical decision-making, our goal is not to guarantee that a decision-maker will reach the "best" possible conclusion. Instead, our purpose is to use this four-step framework to make better decisions in the present and future, justifying certain courses of action in full awareness of the ethical consequences. Every decision-maker will *not* come to the same choice, and every decision will *not* be based on the same rationale. Instead, our hope is to illuminate what is at stake from an ethical perspective to facilitate a better appraisal of the ethical quandaries, as well as provide a path to navigate such terrain.

CONCLUSION

In this chapter, we use ethics as a disciplinary lens in order to clarify issues within current teacher evaluation reform, recognizing that decision-making is far more complicated than simply invoking the adage, "First, do no harm." More specifically, we provide a framework for ethical decision-making in the context of teacher evaluation and accountability demands. We argue that such an approach provides a useful, additive approach for evaluating the impact of SGMs, shedding light on possible considerations for high-stakes educational decision-making.

But is this enough? In a pluralistic democracy comprised of persons with diverging views and assumptions, how can decision-makers be sure that they do not overlook the moral dimensions of their teacher evaluation judgments? In explicating our framework for ethical decision-making, we maintain the complexity of the ethical issues inherent in a hypothetical year of an average educational decision-maker, while also giving decision-makers an approach to help navigate the ethical terrain in teacher evaluation. In other words, the framework provides a context in which to make better decisions in the present and future, justifying certain courses of action in full awareness of their ethical consequences. That context is sensitive to the fact that simple approaches do not exist when dealing with complicated ethical dilemmas. This work is never fully complete; it must be consistently re-engaged as new research emerges and new situations are encountered.

Our approach does not expect that every decision-maker will come to the same choice, or that even if they make a similar choice it will be based

on the same rationale. Instead, our hope is to illuminate the ethical stakes in order to facilitate a better appraisal of the quandaries and provide a path to better navigate the moral dimensions of evaluation. In the current teacher evaluation landscape, especially given the potential impacts associated with the use of SGMs, educational decision-makers cannot afford to ignore these moral dimensions; the consequences are far too severe (Amrein-Beardsley & Collins, 2012; Amrein-Beardsley, 2014; Baker et al., 2013; Lavigne, 2014; Paige, 2012, 2014; Pullin, 2013). We believe this framework for ethical decision-making can help ensure vigilance in regards to intricate issues related to the use of SGMs in teacher evaluation. We offer this framework as a set of steps in the right direction, recognizing the conversation does not, and should not, end here.

Acknowledgments The authors wish to acknowledge helpful comments on an earlier draft of this chapter from Todd DeMitchell, Ph.D., and John Brackett, Ed.D.

REFERENCES

- American Evaluation Association. (2004). *Guiding principles for evaluators*. Washington, DC: American Evaluation Association.
- American Statistical Association. (2014). *Statement on using value-added models for educational assessment*. Alexandria, VA: American Statistical Association. Retrieved from www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf
- Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability*. New York, NY: Routledge.
- Amrein-Beardsley, A., & Collins, C. (2012). The SAS Education Value-Added Assessment System (SAS EVAAS) in the Houston Independent School District (HISD): Intended and unintended consequences. *Education Policy Analysis Archives*, 20(12). Retrieved from <http://epaa.asu.edu/ojs/article/view/1096>
- 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), 1–71.
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., et al. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute. Retrieved from www.epi.org/publication/bp278/
- Beauchamp, T. L., & Childress, J. F. (2001). *Principles of biomedical ethics* (5th ed.). New York, NY: Oxford University Press.

- Berliner, D. C. (2013). Effects of inequality and poverty vs. teachers and schooling on America's youth. *Teachers College Record*, 115(December), 1–26.
- Berliner, D. C. (2014). Exogenous variables and value-added assessments: A fatal flaw. *Teachers College Record*, 116(1), 1–31.
- Bloch, S., & Green, S. A. (2006). An ethical framework for psychiatry. *The British Journal of Psychiatry: The Journal of Mental Science*, 188, 7–12.
- Bunda, M. A. (1985). Alternative systems of ethics and their application to education and evaluation. *Evaluation and Program Planning*, 8, 25–36.
- Collins, C., & Amrein-Beardsley, A. (2014). Putting growth and value-added models on the map: A national overview. *Teachers College Record*, 116(1), 1–9.
- Doherty, K. M., & Jacobs, S. (2015). *State of the states 2015: Evaluating teaching, leading, and learning*. Washington, DC: National Council on Teacher Quality. Retrieved from <http://www.nctq.org/dmsStage/StateofStates2015>
- Drane, J. F. (1982). Ethics and psychotherapy: A philosophical perspective. In M. Rosenbaum (Ed.), *Ethics and values in psychotherapy* (pp. 15–50). New York, NY: The Free Press.
- Dworkin, R. (1979). *Taking rights seriously*. Cambridge, MA: Harvard University Press.
- Elwood, J. (2013). Educational assessment policy and practice: A matter of ethics. *Assessment in Education: Principles, Policy & Practice*, 20, 205–220. doi:10.1080/0969594X.2013.765384.
- Evans, C. M. (2015, June 5). The missing framework: A case for utilizing ethics to evaluate the fairness of educator evaluation systems (Commentary). *Teachers College Record*. Retrieved from <http://www.tcrecord.org>
- Goldhaber, D. (2010). *When the stakes are high, can we rely on value-added? Exploring the use of value-added models to inform teacher workforce decisions*. Washington, DC: Center for American Progress. Retrieved from <http://www.americanprogress.org>
- Gullickson, A. R., & Joint Committee on Standards for Educational Evaluation. (2009). *The personnel evaluation standards: How to assess systems for evaluating educators* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- Harris, D. N. (2011). *Value-added measures in education: What every educator needs to know*. Cambridge, MA: Harvard Education Press.
- Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794–831. doi:10.3102/0002831210387916.
- House, E. R. (1976). Justice in evaluation. In G. V. Glass (Ed.), *Evaluation studies review annual* (Vol. 1, pp. 75–100). Beverly Hills, CA: Sage.
- House, E. R. (1978). Assumptions underlying evaluation models. *Educational Researcher*, 7(3), 4–12.
- House, E. R. (1980). The role of theories of justice in evaluation—Justice on strike. *Educational Theory*, 30(1), 67–72.

- House, E. R. (1993). *Professional evaluation: Social impact and political consequences*. Newbury Park, CA: Sage.
- House, E. R., & Howe, K. R. (1999). *Values in evaluation and social research*. Thousand Oaks, CA: Sage Publications, Inc.
- Howe, K. R., & Moses, M. S. (1999). Ethics in educational research. *Review of Research in Education*, 24, 21–59.
- Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: American Council on Education and Praeger Publishers.
- Kane, M. (2013). Validity and fairness in the testing of individuals. In M. Chatterji (Ed.), *Validity and test use: An international dialogue on educational assessment, accountability and equity* (pp. 17–53). Bingley, UK: Emerald Group Publishing Limited.
- Kitchener, K. S. (1984). Intuition, critical evaluation and ethical principles: The foundation for ethical decisions in counseling psychology. *The Counseling Psychologist*, 12(3), 43–55.
- Kitchener, K. S., & Kitchener, R. F. (2009). Social science research ethics: Historical and philosophical issues. In D. M. Mertens & P. E. Ginsberg (Eds.), *The handbook of social research ethics* (pp. 5–22). Thousand Oaks, CA: Sage Publications, Inc.
- Kitchener, K. S., & Kitchener, R. F. (2012). Ethical foundations. In S. Knapp, M. C. Gottlieb, M. M. Handelsman, & L. VandeCreek (Eds.), *APA handbook of ethics in psychology, Vol 1: Moral foundations and common themes* (1st ed., pp. 3–42). Washington, DC: American Psychological Association.
- Koedel, C., & Betts, J. R. (2009). *Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique*. Cambridge, MA: National Bureau of Economic Research.
- Kymlicka, W. (1990). *Contemporary political theory: An introduction*. New York, NY: Clarendon Press.
- 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.
- Lavigne, A. L., & Good, T. L. (2013). *Teacher and student evaluation: Moving beyond the failure of school reform*. New York, NY: Routledge.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35(11), 1012–1027.
- Messick, S. (1985). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 33–45). Hilldale, NJ: Lawrence Erlbaum Associates, Inc.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: American Council on Education and Macmillan Publishing Company.

- Morris, M. (2003). Ethical considerations in evaluation. In *Springer international handbooks of education: International handbook of educational evaluation*. Retrieved from http://search.credoreference.com/content/entry/spredev/ethical_considerations_in_evaluation/0
- Morris, M. (Ed.) (2008). *Evaluation ethics for best practice: Cases and commentaries*. New York, NY: The Guilford Press.
- Morris, M. (2011). The good, the bad, and the evaluator: 25 years of AJE ethics. *American Journal of Evaluation*, 32(1), 134–151.
- Newman, D. L., & Brown, R. D. (1996). *Applied ethics for program evaluation*. Thousand Oaks, CA: Sage.
- Newton, P. E., & Shaw, S. D. (2014). Twenty-first-century evaluation. In *Validity in educational & psychological assessment* (pp. 183–226). Cambridge, UK: Cambridge Assessment.
- Newton, X. A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Education Policy Analysis Archives*, 18(23), 1–27.
- Nozick, R. (1974). *Anarchy, state, and utopia*. New York, NY: Basic Books.
- Paige, M. A. (2012). Using VAM in high-stakes employment decisions. *Phi Delta Kappan*, 94(3), 29–32.
- Paige, M. A. (2014). A legal argument against the use of VAMs in teacher evaluation. *Teachers College Record*, 166(1), 1–31.
- Pullin, D. (2013). Legal issues in the use of student test scores and value-added models (VAM) to determine educational quality. *Educational Policy Analysis Archives*, 21(6), 1–28.
- Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Belknap Press of Harvard University Press.
- Reardon, S. F., & Raudenbush, S. W. (2009). Assumptions of value-added models for estimating school effects. *Education Finance and Policy*, 4(4), 492–519.
- Ross, W. D. (1930). *The right and the good*. Oxford, UK: Clarendon Press.
- Rothstein, J. (2009). *Student sorting and bias in value added estimation: Selection on observables and unobservables (NBER Working Paper No. 14666)*. Cambridge, MA: National Bureau of Economic Research.
- Scherrer, J. (2011). Measuring teaching using value-added modeling: The imperfect panacea. *NASSP Bulletin*, 95(2), 122–140. doi:10.1177/0192636511410052.
- Schochet, P. Z., & Chiang, H. S. (2012). What are error rates for classifying teacher and school performance using value-added models? *Journal of Educational and Behavioral Statistics*, 38(2), 142–171. doi:10.3102/1076998611432174.
- Scriven, M. (1967). The methodology of evaluation. In *Perspectives of curriculum evaluation* (pp. 39–83). Chicago, IL: Rand McNally & Company.
- Simons, H. (2006). Ethics in evaluation. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *Handbook of evaluation: Policies, programs and practices* (pp. 243–265). Thousand Oaks, CA: Sage.

- Strike, K. A. (1979). The role of theories of justice in evaluation: Why a house not a home. *Educational Theory*, 29(1), 1-9.
- Strike, K. A. (1980). Justice in evaluation: Homecoming rejoinder to house. *Educational Theory*, 30(1), 73-76.
- Strike, K. A. (1990). The ethics of educational evaluation. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers* (pp. 356-373). Newbury Park, CA: Sage.
- Strike, K. A., & Bull, B. (1981). Fairness and the legal context of teacher evaluation. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp. 303-343). Beverly Hills, CA: Sage.
- Stufflebeam, D. L. (1994). Empowerment evaluation, objectivist evaluation, and evaluation standards: Where the future of evaluation should not go and where it needs to go. *American Journal of Evaluation*, 15(3), 321-338.