



Music teachers' perceptions of high stakes teacher evaluation

Mitchell Robinson

To cite this article: Mitchell Robinson (2017): Music teachers' perceptions of high stakes teacher evaluation, Arts Education Policy Review, DOI: [10.1080/10632913.2017.1373380](https://doi.org/10.1080/10632913.2017.1373380)

To link to this article: <https://doi.org/10.1080/10632913.2017.1373380>



Published online: 14 Dec 2017.



Submit your article to this journal [↗](#)



Article views: 76



View related articles [↗](#)



View Crossmark data [↗](#)



Music teachers' perceptions of high stakes teacher evaluation

Mitchell Robinson

School of Music, Michigan State University, Okemos, Michigan, USA

ABSTRACT

Recent corporate education reform policies have replaced relatively informal systems of principal observations that had been familiar to many teachers for much of their professional careers with high-stakes teacher evaluation (HSTE) systems that now determine who is allowed to remain in the profession and who gets terminated. Many education scholars have found current teacher evaluation systems to be lacking in validity and reliability. This article examines the perspectives of music teachers working within HSTE systems through a policy lens, identifying the major challenges these systems pose for music educators and offering policy recommendations for improving the evaluation of music teachers.

KEYWORDS

Education reform; music education policy; teacher evaluation

Introduction

The recent wave of neoliberal, corporate education reform policies has had a strong influence on the nature and extent of teacher evaluation systems. Relatively informal systems of principal observations that had been familiar to many music teachers for much of their professional careers have now been replaced by a highly systematized, data-driven approach that seeks to divide teachers into groups based on their “effectiveness” ratings. Perhaps even more importantly, with tenure protections now absent or weakened in many states, these high-stakes teacher evaluation (HSTE) systems—and the data on which decisions are made—now represent a procedure for determining who is allowed to remain in the profession and who gets terminated.

While state and national policies, such as No Child Left Behind and Race to the Top, have shined a new light on the issue of teacher effectiveness, many educational scholars have found current HSTE systems to be lacking in both validity and reliability. Amrein-Beardsley and Collins (2012) studied the termination of four teachers in 2011 in the Houston Independent School District, and found that these decisions were based largely on the teachers' Value-Added Measures (VAM) scores. A teacher's VAM score is “determined by using a complicated statistical formula that shows whether their students' standardized test scores are higher or lower than predicted” (Holloway-Libell, Amrein-Beardsley, & Collins, 2012, p. 66).

In the case of the four Houston teachers, the researchers found that the VAM scores were likely influenced by

factors not accounted for by the statistical formulas, including changes in student populations, changes in teaching assignments, and student attitudes regarding testing. In another landmark case from New York State, a fourth-grade teacher, Sheri Lederman, sued state education officials over a negative evaluation she received. The judge in the case ruled that the evaluation system, and specifically its use of VAM, resulted in evaluation ratings that were “arbitrary” and “capricious.” Lederman's attorney said that the decision offered “important observations that VAM is biased against teachers at both ends of the spectrum, disproportionate effects of small class size, wholly unexplained swings in growths scores, strict use of curve” (Strauss, 2016).

Organizations from the American Educational Research Association (AERA, 2015) to the American Statistical Association (ASA, 2016) have also come out forcefully against the use of VAMs in high stakes settings, such as teacher evaluation. According to the ASA:

- “VAMs are generally based on standardized test scores and do not directly measure potential teacher contributions toward other student outcomes.
- VAMs typically measure correlation, not causation: Effects—positive or negative—attributed to a teacher may actually be caused by other factors that are not captured in the model” (Strauss, 2014).

Even as the consensus on the use of VAM in high stakes teacher evaluation seems firmly opposed to the practice, the number of states requiring that some type

of “measures of student growth” be a part of their teacher evaluation policy is growing. According to the National Center for Teacher Quality, 43 states “require objective measures of student achievement to be *included in teacher evaluations*” (emphasis mine), 17 states mandate that “student growth (be) the *preponderant* criterion in teacher evaluations” (emphasis mine), and 23 states require that this “evidence of teacher performance *informs tenure decisions*” (emphasis mine) (NCTQ, 2015, p. i). These figures represent a noticeable increase since 2009, when only 15, 4, and 0 states, respectively, held these policy positions.

In spite of the evidence regarding the problems with using student growth measures to determine teachers’ effectiveness ratings and tenure decisions, some states are actually bucking this trend by increasing the percentage of a teacher’s score based on this data. In Michigan, for example, the state legislature passed Public Act 173 in 2015, stipulating the following:

- “The legislation requires that evaluations be conducted annually, and that they incorporate student growth as a significant component, beginning at 25% in the 2015–2016 school year and growing to 40% in 2018–2019. ... Beginning with the 2018–2019 school year, for core content areas in grades and subjects in which state assessments are administered, 50% of student growth must be measured using the state assessments.” (Michigan Department of Education, 2015, pp. 5–7)

This article will consider aspects of current HSTE policy within the context of evidence collected from eight inservice music teachers regarding their experiences with their school districts’ teacher evaluation systems. Teachers’ perspectives were collected through the use of a questionnaire, included in the [Appendix](#), modeled after one used by Ford et al. (2017).

Policy context

For advocates of these teacher evaluation systems, the question of how to motivate teachers to improve their effectiveness is predicated on an assumption that teachers are not intrinsically motivated to do their jobs well, or to the best of their ability. This mindset is based on an employee evaluation model from the business world known as “stack ranking” or “rank & yank” that was popularized by the software giant, Microsoft:

The predominant employee evaluation tool in the business world for years has gone by the name of “stack ranking,” and it should sound familiar to teachers and principals everywhere. The system works by dividing employees into arbitrarily predetermined “effectiveness categories,” based on ratings

by managers. At Microsoft, for instance, employees were rated “on a score of one to five, with one being the best. Managers were then given a curve to base their rankings on, and forced to give a certain percentage of employees a poor ‘five’ label—even if the managers did not consider the employee to be unsatisfactory at their jobs” (Halleck, 2013). This ranking and sorting procedure was used at the software company to determine bonus payments as well as employee terminations, with “a second consecutive low ranking [meaning] that an employee would be terminated” (Halleck, 2013)—the ultimate in high-stakes assessment. (Robinson, 2015, p. 16)

Ironically, even as the business world moves away from this approach to employee evaluation (Brustein, 2013), schools seem to be ramping up their efforts in terms of applying these tenets in the realm of teacher evaluation. Doing so ignores the fact that the vast majority of teachers (Sinclair, 2008), and music teachers in particular (Bergee et al, 2001; Parkes & Jones, 2012), enter the profession for intrinsic, not extrinsic, goals; such as the belief that teachers’ work contributes to the betterment of society, and that their efforts represent a contribution to the common good.

The purpose of teacher evaluation has traditionally been considered as consisting of two elements: pressure and support (Marzano, 2012, p. 14). These characteristics may also be thought of as “accountability” and “the improvement of instruction.” The reform movement’s emphasis on accountability, at the expense of any clear attempts to help teachers use the results of their evaluations to improve their teaching practices (i.e., capacity building), has led to a myopic focus on the “measuring” of teacher effectiveness without much attention to what is actually being measured, what those measurements truly mean, or how these evaluations should be administered across diverse subject areas and teaching contexts.

As Ford et al. (2017) explain:

...for ‘effective’ teachers to persist in the profession, we argue—as others have—that policy must provide both pressure and support (i.e., capacity building) to be successful (Elmore & McLaughlin, 1982; McLaughlin & Pfeiffer, 1988); neither pressure nor support alone will properly motivate those who remain to make desired improvements (McLaughlin, 1987). ... Currently, however, few state policies have articulated plans for how teachers will use formative data in improving their practice, and policy has not been instructive on how the use of value-added information might be effectively used for teacher feedback (Amrein-Beardsley & Collins, 2012; Collins & Amrein-Beardsley, 2014). As with any policy, the design of HSTE policies reflects underlying presuppositions about the degree to which teachers are willing and/or motivated to undertake, with little inducement, the types of improvements to practice that their

evaluations suggest (Firestone, 2014; Schneider & Ingram, 1990). (pp. 205–208)

One way to understand the tensions that may exist with current HSTE systems is to view their actions through the lens of Cognitive Evaluation Theory (CET). CET posits that there are two different kinds of motivation: intrinsic and extrinsic. Intrinsic motivators are things like achievement, recognition of competence, and increased levels of responsibility—characteristics that derive from the satisfaction of an individual doing the tasks associated with one’s work—while extrinsic motivators take the form of compensation, promotions, merit pay, and other inducements—and are controlled primarily by others.

CET was developed to help explain the effects of external inducements on individuals’ intrinsic motivation. The theory suggests that when an employer uses extrinsic targets to induce intrinsically motivated employees to meet the employer’s goals, the result is that the worker’s motivation actually decreases. According to Deci and Ryan (1985), if persons who are motivated by their own internal desire to contribute to a common or greater good believe that they are being expected to work only for the extrinsic rewards of pay, or improved working conditions, their motivation begins to erode. Put another way...

Extrinsic rewards affect intrinsic motivation depending on how recipients interpret them. If recipients believe that the rewards provide positive information about their own competence and self-control over results, intrinsic motivation will increase. If recipients interpret the results as indicating external control, decreasing their feelings of self-control and competence, intrinsic motivation decreases. (Ledford & Fang, 2013, p. 19)

Researchers have found that external reward systems have decidedly mixed results when applied in educational settings. In a landmark study on the effectiveness of merit pay schemes in Massachusetts to entice beginning teachers to remain in the classroom, Liu, Johnson, and Peske (2004) discovered that a \$20,000 signing bonus (paid over a 4-year period) was not enough to motivate their participants to either enter the profession or to stay in their positions for the 4 years required to receive the full bonus—with eight of their 13 participants deciding to leave their positions before that time. The authors suggest that the signing bonus “a. relied too much on inducements and not enough on capacity building, b. focused too narrowly on recruitment and not enough on attention, and c. centered too much on individuals and not enough on schools” (p. 217).

In music education, researchers have examined the phenomenon of arts educator evaluation globally (Shaw,

2016a; West, 2012), and its impact on teacher stress (Shaw, 2016b). This article examines the perspectives of music teachers working within HSTE systems through a policy lens, identifying the major challenges these systems pose for music educators, and offering policy recommendations for improving the evaluation of music teachers.

Empirical support: Teachers’ stories

A total of eight music teachers from five different states were invited to participate in this investigation; seven individuals eventually responded to the questionnaire, and the analysis and interpretation of their responses forms the substance of this article. These teachers were chosen based on their previously expressed interest in and experiences with HSTE in their own school settings, and by their active engagement in their local policy contexts. Due to the locations of these teachers in several states located across the East and Midwest (see Table 1), information was collected primarily via e-mail questionnaire during the summer of 2016, with some follow-up discussions conducted by phone and e-mail over the ensuing months.

The questionnaire (see the Appendix) was modeled after one used by Ford et al. (2017) in their study of teachers in Louisiana, with adjustments made to focus more specifically on the issues encountered by music educators with respect to HSTE. The questionnaire included 11 items, ranging from introductory questions asking each participant to describe their teaching context and setting, to more probing questions addressing specific elements of each teacher’s particular HSTE procedures (i.e., what aspects of the process were creating the most stress and pressure, and what music-specific issues with respect to HSTE were posing any challenges or obstacles).

Teachers’ responses and discussion

It was clear that the pressure and stress associated with HSTE had struck very similar chords in all of the music teachers in the group. The fact that the seven music teachers who so graciously volunteered to assist me in

Table 1. Participants.

Participant	Years of teaching experience	Music teaching specialization area	State
Mr. E	15	High school band	MI
Ms. T	5	Elementary general	KS
Mr. K	33	Elementary band	NY
Ms. Z	17	Middle school band	MI
Mr. S	18	Middle school band	MI
Mr. F	32	Elementary band	NY
Ms. O	6	Middle school chorus	MD

this investigation were willing to share their perspectives so openly and honestly was an indication of their commitment to their profession and how invested they were in their continued growth as teachers.

After multiple “sweeps” of the transcripts I found myself returning to a group of four major ideas that permeated the teachers’ responses:

1. “Fake teaching”
2. Subject-matter inequity
3. “Hoop jumping”
4. Lack of support

The following section describes each of these ideas in more detail and includes quotes from the participants to illustrate their perceptions of each theme.

“Fake teaching”

One of the biggest problems that I faced was my evaluation and the concert schedule. My evaluation went extremely well. I received top marks in almost all areas. ... However, even though it went well, the lesson did not match what I actually teach my students. I used several reading techniques in my teaching to focus more on language aspects of music, which is valuable but only to an extent for me. ... I feel that if I actually did an evaluation the way that I do teach, my administration would not understand. I do use all sorts of techniques and tools, but not to the extent that I demonstrate in an evaluation. It is fake to me so I can keep my job. (Ms. T)

As discussed previously, CET suggests that intrinsically motivated workers do not respond well to external rewards (or threats, as the case may be), and that their feelings of accomplishment and job satisfaction may actually be inhibited when confronted by an evaluation system based on such factors. Struck by the final sentence of Ms. T’s comments above, I came to think of this tension between the expectations of her school district’s HSTE system and her own core beliefs about music teaching and learning as “fake teaching”: a pedagogical game of “Three Card Monte” that created untenable levels of stress and anxiety, manifesting in crippling self-doubt and troubling questions of identity among these teachers.

Tension between HSTE expectations and teachers’ core beliefs

The tension between her district’s expectations and her own beliefs about teaching has been challenging for Ms. T, an elementary general music teacher:

I have to train my students for a few class periods on a different way of teaching so I do have a successful evaluation. I still use developmentally appropriate practices and try to maintain my personal goals, but when I have to integrate many different reading techniques, use of

the word wall (which right now is more for decoration), and more group work, I have to change some things around. ... It has been difficult for me to maintain my core beliefs while, in a way, “teaching to the template.” (Ms. T)

Ms. O, a middle school chorus teacher, expressed her frustration at being forced to administer a test that was incongruous with her own goals for her students:

A frustrating part of the system is the SLO (student learning objective) exam in 6th grade chorus. It is not written in a way that aligns with what or how I am teaching. ... Personally, I do not think our curriculum aligns with this system at all. Unfortunately, it has been more like, “here is how to learn the system and get a good score in it.” It’s awful! (Ms. O)

The anguish in Ms. O’s words is apparent; she knows that what she is being asked to do is not in her students’ best interest, and violates her own beliefs about what constitutes “good music teaching” but feels as though she is between the proverbial “rock and a hard place.”

Not every comment about the changes forced by HSTE expectations was negative. Mr. S, a middle school band teacher, found some of his district’s suggestions quite helpful:

I have altered my instructions by falling into line with what all teachers are required to do. I have “I Can” statements/Target Goals, Word Walls/Word Bank reviews, more use of games to review concepts and performance. I do think that many of these tasks are helpful in keeping my students on task towards our goals. My role has changed from being just a music teacher to a whole student body teacher. (Mr. S)

Self-doubt

Personally the HSTE system unglued me this year. As I mentioned before, I truly believe that I have had one of the most joyous and successful school years to date, yet my score is the lowest it has been and that is extremely frustrating. I am most affected by the scoring aspect and how it makes you question if you are actually effective. I question how it is possible that you could burn yourself out working extremely hard and it still comes down to a number at the end of the year. I am more than a category or a number! (Ms. O)

For Ms. O, the realization that her most successful year of teaching had resulted in a lower rating than she had previously received came as a rude awakening, and caused her to question her effectiveness as a music teacher. The pressures created by the HSTE experience have resulted in crippling bouts of self-doubt for many teachers, leading them to wonder if they have made the right career choice.

Ms. T's struggles negotiating the tension between her beliefs about teaching and the requirements of the HSTE system have led her to question whether she's "good enough" to meet her students' needs:

Personally, I have begun to question whether or not I am a good enough educator. I feel like I have to make a choice between what my students need and meeting the expectations of the district. I don't think there should be a choice. Meeting the needs of my students is the obvious choice, but if I do that, I am not considered a good enough teacher because it does not always fit the rubric. (Ms. T)

When an evaluation system based on extrinsic motivators forces intrinsically motivated workers to doubt their own effectiveness, we must ask ourselves hard questions about the goals of such a system. For teachers to believe that they are "good" at what they do, they must feel a sense of agency for their practice. An evaluative process that chips away at that agency is unlikely to help teachers improve. As we will see below, the absence of any discernible emphasis on capacity building in these evaluation systems has severely limited the role of HSTE to help teachers improve their practice, and to find reasons to persist in the face of crippling self-doubt causing them to question their personal and professional identities.

Questions of identity

Yes, I have questioned my identity as a teacher. When I am administering the pre- and post-tests (whose dubious value and validity I described above), losing instructional and rehearsal time, I feel totally ineffective and part of a "machine" that is doing an injustice to our students. (Mr. K)

For Mr. K, the most experienced teacher in the group with 33 years of service in the public schools, to question his identity as a result of the impact of HSTE on his practice is an indicator of the power that these evaluation systems have over how teachers view their professional lives.

The search for ways to improve their teaching was a common trope throughout the teachers' responses, and reaffirms the notion that most teachers enter the profession for intrinsic reasons, with a commitment to continually improving their work in the classroom with students and colleagues. To do so, however, requires that the feedback with which they are provided, and the information this feedback is based on, is immediate, meaningful, and relevant to what—and who—they are teaching.

Subject-matter inequity

I feel it [his district's HSTE system] has been a completely invalid measurement of my teaching. I teach music. Twenty percent of my evaluation comes from the "building growth score" in two subjects I do not teach. This year I was observed by the Assistant Athletic Director. He had nothing for me other than canned platitudes. My home school principal is an ex-music teacher and music administrator. She is quite thorough and exacting and was an outstanding music teacher and administrator. She knows. She has done at least one of my observations each of the last four years since the dawn of the Annual Professional Performance Review (APPR). Other teachers have principals who know nothing about music teaching as their observer. The inequities are obvious. (Mr. K)

Teacher evaluation systems, in order to assess teachers across a wide range of disciplines, grade levels, and school types, often become so standardized that they reduce teaching down into what are thought to be its component parts. In doing so, these reductionist approaches fail to capture the diversity, variety, and nuance that characterizes good teaching. Nowhere is this failure more pronounced than in the use of generic teacher evaluation systems in the assessment of music teachers.

The teachers here identified five ways in which their HSTE systems created or magnified subject-matter inequities that impacted them as music teachers:

- Invalid measures
- Use of VAM
- "Music teaching stops"
- Lack of control
- Lack of observers fluent in music teaching practices

The following section briefly describes each of these subject-matter inequities, providing quotes from the participants to illustrate their concerns.

Invalid measures

Teachers were frustrated at the poor alignment of HSTE student growth measures and other assessment tools with their subject matter as music teachers, and felt that being forced to use these measures undermined their effectiveness as teachers—and diminished the place of music in the school curriculum. The assessment tools—both music tests and tests of other subjects—these teachers were required to use came from a variety of sources: commercial products, state-mandated tests, and assessments created by local teachers or consortia—but none of them were responsible for designing their own assessments. This lack of agency only exacerbated their frustration at the poor alignment between the mandated assessments and their work as music teachers.

Mr. E, a high school band director, described his discomfort with using measurement tools that do not align with his notions of “good music teaching”: “I thought about the ‘pre’ and ‘post’ test model, but it seems that we desire students to get wrong answers on the pretest, to show growth. It doesn’t sit right with me personally to do that” (Mr. E).

The use of VAM

One of the most controversial features of current HSTE is its reliance on VAM to derive a teacher’s evaluation rating. VAM is especially problematic for music teachers, and teachers of the other “non-tested subjects” that make up roughly 70% of the school curriculum (Watson et al., 2009). These music teachers expressed their frustration at being held accountable for test scores in subjects they did not teach, and from students they had never taught: “I find it remarkably upsetting that my evaluation is affected by reading scores on standardized testing of STUDENTS I HAVE NEVER EVEN MET!” (Ms. Z).

“Music teaching stops”

Several of the teachers remarked on their frustration with infringements on music instructional time for testing in other subjects. They felt that this practice was not only unfair, but sent a disturbing message to students and the community about what was valued in the educational system—and what was not.

At the time of the year where we are building momentum toward Spring Concert season, all teaching and progress slows dramatically while students take the (state) tests. Teaching stops during testing. In one of my buildings, the principal will not allow music instruction in the afternoon on days where the state test is administered in the morning. This practice will not continue next year. (Mr. K)

Lack of control

These music teachers experienced a lack of control over their teaching when HSTE systems appeared to privilege extra-musical objectives and goals (i.e., student test scores in subjects other than music, Lexile reading scores, state-mandated tests) over the teachers’ own goals and the music curriculum in place in the schools. Conversely, these teachers felt the most control over their practice when they were “making music” and when they were working directly with their students in music activities (i.e., selecting repertoire, rehearsing, helping students learn to compose and improvise).

I feel the most control over my teaching in terms of who I am as an individual when it is just my students and me. When we are having fun, exploring music, students are actively participating and asking questions, when we are creating, I feel this is when I am at my best. When I

feel I have the least control over my teaching is when I have to complete an evaluation, when I am in department meetings, when I teach straight from the district curriculum, and when someone else is in the classroom. I feel in these instances, I have expectations to meet and they are not my own. (Ms. T)

Lack of observers fluent in music teaching practices

Teachers had a wide range of responses to being observed by supervisors who did not have much background in or knowledge of music teaching or learning. Ms. Z acknowledged that her supervisors were unlikely to provide her with much in the way of subject-specific feedback, but also saw a “silver lining” in their lack of music background.

I realize that as a music teacher, the likelihood of me ever being evaluated by someone who truly understands what they are seeing is slim to none. I feel fortunate that the three evaluators I have had under this system seem to think that what I do is something akin to wizardry, so I always receive high marks. My administrators don’t understand what I do and would never step in and try to change it. (Ms. Z)

However, the inconsistencies among observers in terms of their levels of music background also posed problems in terms of the perceived usefulness, and even validity, of their evaluations.

Also the rubric for teacher observations is very vague and depends on the experience and personal opinion of the observer, which is why my evaluation scores varied so greatly. The biggest obstacle is being observed by administrators who are unsure what a successful music classroom looks like. (Ms. O)

“Hoop jumping”: Increased expectations of HSTE

The school is constantly expecting more from us, especially in the area of teaching reading and writing, as well as the hoops we need to jump through for evaluations. (Ms. Z)

The expectations placed on these teachers by their HSTE requirements represented extreme demands of time and attention that, in their judgment, could have been used much more productively in other areas of their personal and professional lives. These expectations were noticed primarily in the following three areas:

- Unreasonable and onerous paperwork, documentation, and record keeping requirements
- Time demands
- Increased stress leading to personal problems and health concerns

The following section will provide excerpts from the teachers’ responses that further explain the sources of these frustrations.

Paperwork, documentation, and record keeping

While many general education teachers have class sizes of 20–30 students, it is not unusual for music (and other “special area”) teachers to work with as many as 200, 300, or even more students in the course of a week. This larger “case load” for music teachers creates massive inequities in terms of the data collection and documentation requirements that are common in most HSTE systems:

We are collecting many points of data on every student, and with 220 students and one teacher, this is difficult. Record keeping and the amount of students that we see (is a challenge). There is so much paper work and time just inputting data. (Mr. S)

These teachers were mindful that every minute spent on paperwork was one minute less spent with students, and in an era of high-stakes evaluations, time is among the most precious of commodities.

Time demands

The biggest challenge with this system is the amount of time it takes away from planning and preparation for the classroom. The portfolio alone takes between 10–12 hours if you do it correctly. The amount of time I spend tracking all of the things I am doing is sickening! (Ms. O)

The increase in time demands associated with their HSTE responsibilities created a major problem for this group of teachers. When one considers the heightened administrative duties concomitant with many music teaching positions (i.e., budget and travel requests, instrument maintenance and repair, uniform fittings, music library records, booster and parent club meetings, football games, parades, musicals, concerts, informances, recitals, etc.), the time management tasks required by HSTE systems pose daunting challenges for music teachers.

Increased stress leading to personal problems and health concerns

Stresswise, I’d have to say that on a scale of 1–10, with a 10 being “please set my hair on fire now,” the 2015–16 school year was about an 11 for me. (Ms. Z)

Based on my observations of teachers over the last decade or so, and considering the explosion in stress created by additional workplace demands and increased job expectations, I do not think it is an exaggeration to say that we are facing a public health crisis in today’s teaching force. The findings from a 2016 issue brief from the Robert Wood Johnson Foundation agree with this assertion, and offer the following synopsis:

Today, teaching is one of the most stressful occupations in the U.S. High levels of stress are affecting teacher health and well-being, causing teacher burnout, lack of engagement, job dissatisfaction, poor performance, and some of the highest turnover rates ever. Stress not only has negative consequences for teachers, it also results in lower achievement for students and higher costs for schools. A New York City study showed higher teacher turnover led to lower fourth and fifth grade student achievement in both math and language arts. The cost of teacher turnover is estimated to be over \$7 billion per year. (Greenberg et al., 2016, p. 2)

These music teachers identified a number of stress-induced problems they believed were the result of the heightened expectations and workload associated with the HSTE systems in their school districts, including lack of sleep, weight gain, less time for spouses, partners, children, and a need for therapy to help deal with these concerns. Teachers sought assistance in dealing with the stresses caused by HSTE in a variety of ways, including asking physicians for anxiety medication, seeking therapy, and “venting” to family members and colleagues. All of the teachers worried about the effects of these increased pressures on colleagues who did not have access to these support systems, or who were not “brave” enough to ask for help.

If I was apprehensive about asking for help I think I would not be in a great place emotionally and mentally. I also see a therapist and that is helpful. The support that I reached out for has been enough to keep my head above the water but I feel concern for those who are not brave enough to reach out! (Ms. O)

Lack of support

I honestly am unsure [of what his district was doing to support teachers with respect to HSTE]—I haven’t been approached or briefed about the situation for the upcoming school year. This may be disseminated to individual departments, but nothing has been brought up to me about curriculum changes. (Mr. E)

Mr. E and the other teachers were extremely frustrated at the lack of support they received from their district and administration. They were anxious for advice and guidance, yet received very little in the way of useful information that could be used to meet the demands and expectations placed on them by the HSTE requirements in their setting. In some cases this meant that teachers simply “fit in” or “taught to the test” in order to fill out the required paperwork and documentation—creating even more stress and role confusion.

Ford et al. (2017) identified the absence of teacher “buy-in” and the almost constant moving of targets and expectations with respect to HSTE as major problems in terms of its successful and effective implementation:

First, with any new innovation, teacher support for change and development, in all its forms (autonomy, competence, and relatedness), needs to be present and focused. Movement from extrinsically based policies to those which build upon and further cultivate teachers' innate intrinsic motivation for teaching and learning would constitute nothing less than a wholesale paradigm shift in thinking about the purposes of teacher evaluation.

Second, though Louisiana “eased” into their implementation of *Compass* and *CCSS* [Common Core State Standards] by implementing one in each of the past 2 years, our evidence suggests that even this is too quick a time frame, particularly when the consequences of poor teacher performance are being enforced. (p. 234)

A troubling addendum...

Given that the ostensible goal of HSTE policies is to weed/counsel out ineffective teachers, it seems as though the policy, at best, is not having the intended effect—quite the opposite in fact. (Ford et al., 2017, p. 234)

As I read through these music teachers' comments I was struck by the realization that there was almost no connection between each individual's expressed feelings of self-efficacy and the ratings they received on their teacher evaluations. Indeed, every one of these persons had consistently received ratings of either “Effective” or “Highly Effective” over their careers, ratings that represented the two highest effectiveness rankings possible on the HSTE rubrics—and still, the teacher evaluation systems being used in their schools had caused them to question their effectiveness in the classroom, doubt their continuation in the profession, and question their professional and personal identities.

This finding was also corroborated by Ford et al. (2017):

Of course, the overall feelings of teachers in our sample by the end of their second year with these initiatives are, in and of themselves, troubling. Equally troubling is the fact that these feelings and experiences coincided with either an “effective proficient” or “highly effective” rating on their overall evaluation—the two highest ratings on the *Compass* scale. An important question to be asked is, “If teachers performed so well on their *Compass* evaluation, why aren't they happier and more committed to teaching?” In addressing this question, two points are worth mentioning. First, our evidence suggests that “good marks” on evaluations seemed not to be what was driving teachers' feelings of self-efficacy, satisfaction, and professional commitment in our sample; instead it was their seeking out of the intrinsic rewards of teaching as well as the feeling that their autonomous actions were driving results. (pp. 232–233)

The reasons for these teachers' seemingly incongruous feelings may have had as much to do with their

personality types and goal orientation structures as with the HSTE rubrics. Many of the participants recognized they were “wired,” so to speak, to accept nothing less than the very highest ratings, regardless of the task.

It was clear that these teachers struggled with balancing the expectations of their schools' HSTE requirements with their own notions of what it meant to be an effective music teacher. These struggles created tensions that manifested in diminished feelings of self-efficacy, classroom control, and decreases in personal and professional worth and value. The following section of this article offers suggestions for policy initiatives to address these issues.

Policy implications

For music teachers who enter their profession for largely intrinsic reasons, being forced to comply with an evaluation system based on primarily extrinsic motivational beliefs creates major obstacles and challenges to their feelings of self-efficacy, professional and personal identity, and even continuation in the profession. This was clearly the case for this group of music teachers—and they are not alone.

A 2015 survey of nearly 3,000 teachers conducted by the Network for Public Education came to the following conclusions:

- Teachers and principals do not believe that evaluations based on VAM scores are valid or reliable measures of their work.
- The emphasis on improving test scores has overwhelmed every aspect of teachers' work
- Teachers report evidence of bias against veteran educators, contributing to a decline in teachers of color, veteran teachers, and those serving students in poverty
- Professional development tied to HSTE is undermining teacher autonomy and limiting their capacity for real professional growth

If we are truly interested in designing teacher evaluation systems that are intended to build capacity for growth among teachers, and are aligned with teachers' actual core beliefs and motivations, there must be an increased emphasis on using the results of these evaluation tools to help teachers improve their practice while at the same time eliminating the punitive elements of the current system's focus on “accountability.”

Recommendations

Provide discipline-specific professional development support for music teachers in the HSTE process

These music teachers identified a lack of support from their administrators, school districts, and state education

departments as they attempted to understand the expectations and requirements of their HSTE systems. The “subject inequity” factors associated with being a teacher of an “untested subject” in a “one size fits all” evaluation system only exacerbated the severity of these issues. So, what does better professional development support for music teachers look like?

First, it is discipline-specific: that is, music teachers require professional development that is related to their work as music teachers, not “generic” support, or sessions designed to address the needs of teachers of other subjects. Second, it is something that is designed *by* music teachers, not done to them. Professional development is a process engaged in and created by its participants; not a product administered *to* subjects. Third, it involves creating “communities of practice” that provide music teachers with opportunities for common planning time, collaborative working arrangements, and team teaching. Fourth, it makes space for music-making (i.e., instrument/voice/conducting lessons; performing with a community band, orchestra, or choir; forming chamber ensembles, garage bands, folk music groups for personal enjoyment and musical fulfillment) as a valid and acceptable form of professional development. Fifth, it considers the pursuit of continued education and graduate study as professional development; and finally, it validates “action” or “teacher research” as professional development (Conway & Edgar, 2014).

Eliminate the use of VAM in music teacher evaluation

The use of student test scores to evaluate teachers has been characterized as an invalid and unsupportable practice by professional associations from AERA to ASA, and its use in determining teacher effectiveness has been termed as “arbitrary” and “capricious” by a New York State Supreme Court justice. Nonetheless, most HSTE systems persist in using VAM in calculating teacher evaluation scores. The music teachers here also identified the use of VAM in their evaluations as a frustrating aspect of the HSTE process: “I was rated Highly Effective this year, pending the results of some State testing taken by students I don’t see in subjects that I don’t teach (that will be included in my final score)” (Mr. F).

Return the focus of teacher evaluation to capacity building and improvement of instructional practices, and, accordingly, change evaluation reporting protocols from numerical to narrative

Teaching, like any profession, is complicated and complex in nature. Teachers work in an astonishingly wide and diverse array of subjects, grade levels, and school settings and contexts. The notion that all teachers can be effectively evaluated on the same four-point scale is

simultaneously naïve and offensive. When the dual purposes of evaluation (i.e., accountability and the improvement of instruction) are uncoupled, the result becomes focused solely on punishment, while ignoring capacity building—and represents a perversion of the educational process. If we are serious about designing an evaluation system that has at its core the elevation of teaching practice, we should consider the process of teacher evaluation one of continuous improvement through rich, reflective conversations between colleagues—not as a matter of checking boxes and jumping through hoops to arrive at a number between 1 and 4. As Ms. T said, “I question how it is possible that you could burn yourself out working extremely hard and it still comes down to a number at the end of the year. I am more than a category or a number!”

Reduce the amount of paperwork, documentation, and record keeping required of music teachers—and administrators—by HSTE procedures

The music teachers here identified the overwhelming paperwork and documentation requirements of their HSTE systems as one of the worst aspects of the process. Mr. S talked about the impact that these demands had on his teaching:

Record keeping and the amount of students that we see. The music staff is allowed to meet a few times a year to analyze data and work on instruction and share ideas. The data and analyzing of data has been extremely time consuming. (Mr. S)

State and district teacher evaluation procedures should be redesigned so as to require significantly less time devoted by teachers to filling out forms and providing “evidence” of teaching effectiveness, allowing teachers to use that time in the classroom, working with students and colleagues. School administrators would also benefit from more streamlined reporting procedures given the often unreasonable observation and documentation demands created by these systems.

(Re?)Align the goals of music teacher evaluation with the professional knowledge base (i.e., research findings and “best practices”) in music education

A consistent theme among this group of teachers was the “disconnect” between their core beliefs about music teaching and learning and the expectations of the HSTE system that dictated their evaluation rating. Labeled as “fake teaching” in this article, these tensions created serious and significant problems for these teachers as they sought to reconcile their daily actions with the requirements posed by an evaluation process that seemingly

ignored their true motivations for becoming teachers, and was ignorant of their practice as music teachers.

Aligning these two currently dichotomous systems will require extensive examination of what constitutes “good music teaching” across diverse contexts, what constitutes culturally relevant music education for diverse learners, the role of music in the school curriculum, and the place of music in society. That we have rushed to matters of ranking and sorting teachers before engaging in these difficult discussions should tell us a great deal about the true motivations of our current HSTE systems.

A good start would be the recognition that music teacher evaluation should become less “standardized” and more nuanced, that “good music teaching” in one place may not look like “good music teaching” in another, and that for teacher evaluation to be meaningful to teachers they must be involved in its design—not just as the unwilling recipients of its administration. These policy changes will require that those closest to the teaching/learning process (i.e., students, teachers, and families) be granted more control over the teacher evaluation process. To do so will require careful and continued interrogation of the policy levers that have been engaged to create the current system.

References

- AERA. (2015). Statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, 44(8), 448–452. doi: 10.3102/0013189X15618385
- Aguilar, C. E., & Richerme, L. K. (2014). What is everyone saying about teacher evaluation? Framing the intended and inadvertent causes and consequences of race to the top. *Arts Education Policy Review*, 115(4), 110. doi:10.1080/10632913.2014.947908
- 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. doi:https://doi.org/10.14507/epaa.v20n12.2012
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W. H. Freeman.
- Bandura, A. (1977). *Social learning theory*. New York, NY: General Learning Press.
- Bandura, A., & Walters, R. (1963). *Social learning and personality development*. Chicago, IL: Holt, Rinehart and Winston, Inc.
- Bergee, M. J., Coffman, D. D., Demorest, S. M., Humphreys, J. T., & Thornton, L. P. (2001). Influences on collegiate students' decision to become a music educator. Retrieved from <http://www.menc.org/resources/view/influences-on-collegiate-students-decision-to-become-a-music-educator>
- Brand, M. (2009). Exhausted from educational reform. *Bulletin of the Council for Research in Music Education*, 180, 87–92.
- Brustein, J. (2013, November 13). Microsoft kills its hated stack rankings. Does anyone do employee reviews right? Bloomberg Businessweek. Retrieved from <http://mobile.businessweek.com/articles/2013-11-13/microsoft-kills-its-hated-stack-rankings-dot-does-anyonedo-employee-reviews-right>.
- Collins, C., & Amrein-Beardsley, A. (2014). Putting growth and value-added models on the map: A national overview. *Teachers College Record*, 116(1). Retrieved from <http://www.tcrecord.org/Content.asp?ContentId=17291>
- Conway, C., & Edgar, S. (2014). Inservice music teacher professional development. In C. Conway (Ed.), *The Oxford handbook of qualitative research in American music education* (pp. 479–500). New York, NY: Oxford University Press. doi:10.1093/oxfordhb/9780199844272.013.025
- Deci, E., & Ryan, R. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum Press.
- Denzin, N. K. (2001). *Interpretive interactionism* (2nd ed.). Thousand Oaks, CA: Sage.
- Ledford, G. E., & Fang, M. (2013). Negative effects of extrinsic rewards on intrinsic motivation: More smoke than fire. *World at Work Journal*, 22(2). Retrieved from <https://www.worldatwork.org/waw/worldatworkjournal/html/journal-2ndQ13.html>
- Elmore, R. F., & McLaughlin, M. (1982). Strategic choice in federal education policy: The compliance-assistance trade-off. In A. Lieberman & M. W. McLaughlin (Eds.), *Policy-making in education, 81st yearbook of the National Society for the Study of Education* (pp. 159–194). Chicago, IL: University of Chicago Press.
- Firestone, W. A. (2014). Teacher evaluation policy and conflicting theories of motivation. *Educational Researcher*, 43, 100–107. doi:10.3102/0013189X14521864
- Ford, T. G., Van Sickle, M. E., Clark, L. V., Fazio-Brunson, M., & Schween, D. C. (2017). Teacher self-efficacy, professional commitment, and high-stakes teacher evaluation policy in Louisiana. *Educational Policy*, 31(2), 202–248. doi:10.1177/0895904815586855
- Glaser, B. S. (1965). The constant comparative method of qualitative analysis. *Social Problems*, 12(4), 436–445. doi:10.2307/798843
- Greenberg, M. T., Brown, J. L., & Abenavoli, R. M. (2016). *Teacher stress and health effects on teachers, students, and schools*. University Park, PA: Edna Bennett Pierce Prevention Research Center, Pennsylvania State University. Retrieved from http://www.rwjf.org/content/dam/farm/reports/issue_briefs/2016/rwjf430428
- Halleck, M. (2013, November 13). What is stack ranking? Microsoft ends controversial employee rating system, Yahoo ramps it up. *International Business Times*. Retrieved from <http://www.ibtimes.com/what-stack-ranking-microsoft-ends-controversial-employee-ratingsystem-yahoo-ramps-it-1468850>
- Hargreaves, D. J., Purves, R. M., Welch, G. F., & Marshall, M. A. (2007). Developing identities and attitudes in musicians and classroom teachers. *British Journal of Educational Psychology*, 77, 665–682. doi:10.1348/000709906X154676
- Holloway-Libell, J., Amrein-Beardsley, A., & Collins, C. (2012). All hat and no cattle. *Educational Leadership*, 70(3), 65–68. Retrieved from <http://www.ascd.org/publications/educational-leadership/nov12/vol70/num03/All-Hat-and-No-Cattle.aspx>

- Instant News West U. (2011). HISD Board Of Education approves eliminating 950 teachers' positions. Retrieved from <http://www.instantnewswestu.com/hisd-board-of-education-approves-eliminating-950-teachers-positions/>
- Liu, E., Johnson, S. M., & Peske, H. G. (2004). New teachers and the Massachusetts signing bonus: The limits of inducements. *Educational Evaluation and Policy Analysis*, 3(26), 217–236. doi: 10.3102/01623737026003217
- Marzano, R. (2012). The two purposes of teacher evaluation. *Educational Leadership*, 70(3), 14–19. Retrieved from <http://www.ascd.org/publications/educational-leadership/nov12/vol70/num03/The-Two-Purposes-of-Teacher-Evaluation.aspx>
- McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis*, 9, 171–178.
- McLaughlin, M. W., & Pfeiffer, R. S. (1988). *Teacher evaluation: Improvement, accountability, and effective learning*. New York, NY: Teachers College Press.
- Michigan Department of Education. (2015). Michigan educator evaluations at-a-glance. Retrieved from http://www.michigan.gov/documents/mde/Educator_Evaluation_s_At-A-Glance_522133_7.pdf
- National Center for Teacher Quality. (2015). *State of the States: Evaluating Teaching, Leading and Learning. Report*. Retrieved from <http://www.nctq.org/dmsView/StateofStates2015>
- Network for Public Education. (2016). *Teachers Talk Back: Educators on the Impact of Teacher Evaluation*. Retrieved from <https://networkforpubliceducation.org/wp-content/uploads/2016/04/NPETeacherEvalReport.pdf>
- Parkes, K., & Jones, B. (2012). Motivational constructs influencing undergraduate students' choices to become classroom music teachers or music performers. *Journal of Research in Music Education*, 60(1), 101–123. doi:10.1177/0022429411435512
- Robinson, M. (2015). The inchworm and the nightingale: On the (mis)use of data in music teacher evaluation. *Arts Education Policy Review*, 116(1), 9–21. doi:10.1080/10632913.2014.944966
- Royo, J. L. (2014). Self-efficacy in music education vocal instruction: A collective case study of four undergraduate vocal music education majors. (Unpublished doctoral dissertation). University of Arizona, Tucson, AZ. Retrieved from <http://arizona.openrepository.com/arizona/handle/10150/320010>
- Scheib, J. W. (2003). Role stress in the professional life of the school music teacher: A collective case study. *Journal of Research in Music Education*, 51(2), 124–136. doi:10.2307/3345846
- Schneider, A., & Ingram, H. (1990). Behavioral assumptions of policy tools. *The Journal of Politics*, 52(2), 510–529. Retrieved from <http://www.jstor.org/stable/2131904>
- Seibt, R., Spitzer, S., Druschke, D., Scheuch, K., & Hinz, A. (2013). Predictors of mental health in female teachers. *International Journal of Occupational Medicine and Environmental Health*, 26(6), 856–869. doi:10.2478/s13382-013-0161-8
- Shaw, R. D. (2016). Arts teacher evaluation: How did we get here? *Arts Education Policy Review*, 117(1), 1–12. doi:10.1080/10632913.2014.992083
- Shaw, R. D. (2016b). Music teacher stress in the era of accountability. *Arts Education Policy Review*, 117(2), 104–116. doi:10.1080/10632913.2015.1005325
- Sinclair, C. (2008). Initial and changing student teacher motivation and commitment to teaching. *Asia-Pacific Journal of Teacher Education*, 36(2), 79–104. doi:10.1080/13598660801971658
- Strauss, V. (2016). Judge calls evaluation of N.Y. teacher “arbitrary” and “capricious” in case against new U.S. secretary of education. Retrieved from https://www.washingtonpost.com/news/answer-sheet/wp/2016/05/10/judge-calls-evaluation-of-n-y-teacher-arbitrary-and-capricious-in-case-against-new-u-s-secretary-of-education/?utm_term=.0f1468aa9c11
- Strauss, V. (2014). Statisticians slam popular teacher evaluation method. Retrieved from https://www.washingtonpost.com/news/answer-sheet/wp/2014/04/13/statisticians-slam-popular-teacher-evaluation-method/?utm_term=.604637c1f223
- Wangberg, E. (1987). Listening to teachers: The missing link in reform. *The Clearing House*, 61(2), 76–80. doi:10.1080/00098655.1987.11478572
- Watson, J. G., Kraemer, S. B., & Thorn, C. A. (2009). *The other 69 percent*. Center for Educator Compensation Reform. U.S. Department of Education, Office of Elementary and Secondary Education, Washington, D.C. <http://smte.us/wp-content/uploads/2011/11/Performance-of-Teachers-of-Non-tested-Subjects-and-Grades.pdf>
- West, C. (2012). Teaching music in an era of high-stakes testing and budget reductions. *Arts Education Policy Review*, 113(2), 75–79. doi:10.1080/10632913.2012.656503

Appendix: Questionnaire

1. TELL ME ABOUT YOURSELF: Please describe briefly your background in education and music and your current responsibilities: Years of experience teaching, degrees earned/institutions attended, teaching certification, grade level(s) taught, music teaching specializations/expertise, outside school-related music involvement, professional associations and memberships.
2. OVERALL BACKGROUND: Tell me how this school year is going for you. What are some successes? Some challenges?
3. WHAT HAS IT BEEN LIKE ADJUSTING TO THE TEACHER EVALUATION (HSTE) SYSTEM IN USE AT YOUR SCHOOL? Tell me about the teacher evaluation system at your school/district. What has been good about the system? Where have the challenges been? Probes: Opinions/perceptions about the idea of HSTE; impact on students; developmentally appropriate practices; training/preparation/ professional development (PD).
4. HSTE SPECIFICS: Are there specific elements of the HSTE system being used at your school that cause more frustration/distress than others? Which ones and why? Probes: SLOs, specific aspects of the HSTE rubric used at your school.

5. DISTRICT/SCHOOL CONTEXT: Now that the HSTE system is being fully implemented in your state, what has your school/district done differently this year to incorporate this approach into your curriculum?
6. MUSIC SPECIFIC ISSUES: What specific issues with respect to HSTE for music teachers, because of our content area and what we teach, pose any challenges or obstacles for you? Probes: Has instructional time for music in your school been affected as a result of HSTE policies, and if so, how? Has the district's formal curriculum for music changed since HSTE systems went into effect, and if so, how? In what ways have you as a music teacher altered your classroom practices as a result of HSTE, and were these teaching strategies effective? In what ways have school music programs and your responsibilities as a music teacher changed since HSTE went into effect?
7. HOW ARE YOU HANDLING HSTE? What is the aspect of HSTE that you feel has affected you the most? Do you feel as though there is a system in place to support you in addressing this area of concern? From where or from whom are you getting this support? What other forms of support would you find useful or helpful with respect to your evaluation? How do you think the HSTE system has affected your physical, mental, and/or emotional health? How have you handled any increased demands, stresses or pressures that the HSTE system has created for you as a teacher?
8. HOW DO YOU FEEL ABOUT HSTE? With respect to HSTE, what have been your initial impressions of this evaluation tool? Do you see any benefits to this evaluation system? What are the drawbacks of such an evaluation system in your opinion? Probes: Student Learning Objectives (SLOs); Value-Added Measures; observation rubric and scaling; training and support.
9. EFFICACY EXAMPLES: Can you give an example of when you feel the most control over your teaching in terms of who you are as an individual and a professional? As a musician and conductor/teacher? Can you give an example of when you feel the least control over your teaching?
10. IDENTITY AND EFFICACY: Since the beginning of this HSTE system, have you ever found yourself questioning your identity either professionally or personally? Your effectiveness? Would you mind describing a good example of one of these instances?
11. HSTE FOLLOW-UP: Can you tell me about your initial reaction to your overall HSTE score? How are you dealing with your rating personally? How well do you feel your evaluation score reflects your teaching? Will you share your overall rating with me?