

# Scaling Up SSTEP in Physical Education Teacher Education: Possibilities and Precautions

Michael W. Metzler

## Introduction

While research on teaching and learning in teacher education programs has a long and impressive history, its offspring, the self-study of teacher education practice (SSTEP)<sup>1</sup> is still very much in the early stages of development. If the establishment of a dedicated, refereed, scholarly journal can be regarded as a key indicator of a field's emergence, then SSTEP reached that distinction just 9 years ago with the inaugural publication of *Studying Teacher Education: A journal of self-study of teacher education practices* in 2005. Certainly, some SSTEP inquiry was conducted well before then (Loughran 2005), but SSTEP research was typically not the main content of any of the journals and books in which this research was published. Other indicators have emerged, such as the establishment of a SSTEP Special Interest Group (SIG) in the American Educational Research Association, and its 'Castle Conference', but again, those developments have been recent.

In that inaugural issue of *Studying Teacher Education*, Editor John Loughran stated that '...self-study has emerged from and been influenced by a range of events and has been built on the works of fields such as reflective practice, action research, and practitioner research' (Loughran 2005, p. 5). To this day, those modes of inquiry have been predominant in *Studying Teacher Education* and represented in SSTEP as a field. They can be characterized, but certainly not stereotyped, as small-scale, short-duration, qualitative studies of mostly individual teacher educators, some of their students, and limited parts of their programs. Zeichner (2007) adds to this

---

<sup>1</sup>The author acknowledges that other terms and conceptualizations of SSTEP have emerged in this line of scholarship, such as *self-study research* (Zeichner 2007). The term SSTEP, which specifically includes the *practices* of teacher education was adopted early by the GSU PETE faculty to guide the research project described in this chapter.

M.W. Metzler (✉)

Georgia State University, Atlanta, GA, USA

e-mail: [mmetzler@gsu.edu](mailto:mmetzler@gsu.edu)

characterization by commenting that few SSTEP studies have been conducted within larger research programs or connected to other lines of research in teacher education. These studies have served the nascent field of SSTEP well and should continue as integral inquiry going forward. However, the question is raised here: Can SSTEP inquiry be scaled up to conduct studies that involve multiple investigators, entire cohorts of students, a broad range of program components, with longitudinal and mixed methods designs? If realized, would that shift the foundation of SSTEP inquiry as we now know it, or would it provide additional ways to understand the practice of teacher education? It should be made clear that scaling-up SSTEP is not meant to set aside the existing methodological traditions; it is proposed as a means to arrive at understandings of teacher education that I believe are not attainable with the current scope and modes of inquiry in SSTEP. In this chapter I will describe a longitudinal self-study of the PETE program at GSU. That project will be used to highlight both the possibilities and some necessary precautions for designing and conducting large-scale SSTEP in PETE.

### *Defining ‘Self’ and ‘Practice’ in SSTEP*

While not exclusive, the primary participants in the practice of teacher education are professors and students; in some circumstances and at certain times, the participation of P-12 teachers is also primary. Attributable in large part to the predominant traditions of inquiry in SSTEP (i.e., reflective practice, action research, and practitioner research) it can be argued that the ‘self’ in SSTEP is more often an individual teacher educator who is attempting to explore and communicate his/her own personal experience and/or meaning as a teacher educator to an audience (Loughran 2010; Zeichner 2007).

In the GSU SSTEP project that is described in the next section, we came to define ‘self’ from a collective perspective that included all faculty members who held major and regular instructional and supervisory responsibilities, involved graduate students, and PETE collaborators from other universities. Such an approach is similar to the notion of ‘institutional self’ as described by Hamilton (2002) and Loughran (2010). Clearly, we all learned individually from our SSTEP work, but we valued and pursued collaborative effort and collective learning first. From that commitment, we have used our institutional SSTEP data as the source for many and often extended group discussions about what those data are telling us, and what we as a group should do with that knowledge to improve our program.

Similarly, we have expanded our functional definition of ‘practice’ in SSTEP to include not only descriptions of what we do to design, implement, and manage the curriculum, but also our efforts to study the impact of those practices on the program itself. Studying program effectiveness and the impact of decisions made at the program level has been an integral part of our SSTEP philosophy and project design from the very beginning (Metzler and Tjeerdsma 1998, 2000).

## The GSU PETE Assessment Project

The GSU Physical Education Teacher Education Assessment Project (PETEAP) began in the 1994–1995 academic year and has been ongoing since then. Initially the purpose of PETEAP was to compare students in the pre-1994 program with cohorts in the post-1995 program on many measures of teacher content knowledge, pedagogical content knowledge, and dispositions. That purpose proved to be unattainable due to the lack of sufficient numbers of pre-1994 students to use for comparisons with later cohorts. We then re-purposed the project to examine the program's effectiveness at meeting its major outcomes for our pre-service teachers. By faculty consensus, those outcomes were identified as the National Association for Sport and Physical Education's (NASPE) Standards for Beginning Physical Education Teachers (NASPE 1995), which were coincidentally established as the PETEAP began. In addition to studying program effectiveness, we also designed mechanisms for studying the teacher education practices that were used to pursue those standards in the program.

Galluzo and Craig (1990) suggest that the place of program assessment research (and similarly SSTEP) can be elevated simply by being more clear about why a teacher education faculty/department would take on assessment in the first place. Once the main purpose of such efforts becomes clear, the faculty can better determine a plan for doing assessment. Galluzo and Craig (1990) propose four main purposes for teacher education program research and assessment: (1) Accountability – to meet external accreditation review standards; (2) Improvement – to gather and use data for making program revision/improvement decisions; (3) Understanding – to understand the experiences of pre-service teachers in the program; and (4) Knowledge – to increase the existing body of knowledge on teacher education – to generalize.

While all four of these purposes can drive program assessment and research, Galluzo and Craig (1990) advocate that the 'overarching purpose' should be to 'develop a comprehensive knowledge-production effort about the relationships among a program's context, inputs, processes, and products' (p. 606). It is clear that the improvement purpose should be placed ahead of the others whenever possible. We agreed with that at the start of the PETEAP, and have maintained that priority throughout the entirety of the project (Metzler and Tjeerdsma 1998). Improvement is determined from multiple sources of data, such as interviews and surveys of program completers, interviews and surveys of cooperating teachers, document analyses, direct observations of instructional practice, measures of teaching efficacy, and more.

During the early years of PETEAP we designed data collection methods and protocols to conduct SSTEP analyses to address questions relative to the initial Standards for Beginning Physical Education Teachers (NASPE 1995), as well as additional questions based on program-specific content knowledge and pedagogical content knowledge (e.g., MBI). Our data gathering efforts quickly grew to include a large number of data sources (students, faculty, P-12 pupils), methods (surveys, interviews, direct observations, and more), and administrative matters (pre-admission,

**Table 1** The GSU PETEAP data collection matrix

	Pre-admission	In “Benchmark” courses <sup>b</sup>	Start PCK sequence <sup>c</sup>	End PCK sequence	During student teaching	End student teaching/ program <sup>d</sup>
Teacher efficacy scale	X		X	X	X	X
Biodata and interviews	X					
Teacher/coach warrant and grade preference <sup>a</sup>	X			X		X
Assignments, projects, exams		X	X	X	X	
Analysis of instructional planning, skills, and assessment			X	X	X	X
Analysis of model-based instruction				X	X	X
Students’ assessment of program						X
P-12 pupils’ perceptions of teaching effectiveness					X	
Program coherence		X	X	X	X	X

<sup>a</sup>Expressed preferences for teaching v. coaching responsibilities and grade levels to teach

<sup>b</sup>Movement and skill content courses in the program

<sup>c</sup>The pre-student teaching practicum courses that develop pedagogical content knowledge

<sup>d</sup>The end of student teaching and the completion of the program occur at the same time

in key courses, pre-student teaching, during student teaching, and after student teaching [end of program]). This plan is summarized in the PETEAP data collection matrix shown in Table 1.

By 2006 we had a well-established plan that annually produced a prolific amount of data and artifacts used to answer many questions about our program. It was also around that time that all teacher education programs in the United States started to be held more accountable for generating and reporting data to external agencies. The list of those agencies grew to include: several administrative units on campus, state teacher licensing commissions, NASPE, program accrediting organizations, state legislative bodies, and more. On first glance it appeared that most of these agencies were requesting data that were similar to what we were already producing and using for our institutional SSTEP. However, these agencies imposed more and

varied data-reporting requirements, making it clear that those added requirements would place a large burden on our time and personnel resources for SSTEP. So, 2006 started a watershed period when we began to shift in how we approached our SSTEP, including elements of our original faculty-designed SSTEP *and* elements that were externally driven. More on that shift will be discussed later in this chapter.

The PETEAP also features a longitudinal approach to SSTEP. We have collected, stored, and analyzed data and artifacts on nearly all of the 350 students who have entered the program since 1994. In addition, we have identified 18 cohorts of students who began and completed the program since that time. A cohort is formed in the term an intact group starts the program and completes it approximately 3 years later. With this design we can study trends over long periods of time, and we can compare cohorts of students who completed the program before certain major changes were made with cohorts of students who completed the program after those changes were implemented. The expanded scope of the PETEAP design has allowed us to conduct longitudinal, large-scale studies of the pre-service PETE program and its graduates (Metzler and Tjeerdsma 1998, 2000). Because many of our graduates have remained in the Atlanta Metropolitan area to teach, we have also been able to study their successes and challenges as they attempted to use MBI during and beyond their professional induction years (Gurvitch and Blankenship 2008; Gurvitch et al. 2008).

We also made another key commitment as part of our early SSTEP efforts – to use instrumentation and protocols that could generate ‘research quality’ data, which can be differentiated from ‘assessment quality’ data (Metzler and Tjeerdsma 1998, 2000). The former meant that we would seek out and use data collection and analytic methods that would meet standards for acceptance in refereed research publications as much as possible. The latter meant data that would be collected to meet less-stringent collection and reporting requirements needed for the previously mentioned external agencies. What we discovered is that by meeting the standards for research quality data as often as possible, we were avoiding duplicative work needed to meet the other reporting requirements; research quality data could be used for assessment reports, but assessment quality data could not be used for scholarly research reports.

## **Possibilities with Longitudinal, Large-Scale SSTEP**

Current SSTEP scholarship might be fairly described as predominantly: individual, introspective, practice-oriented, and short-term (Zeichner 2007). That characterization is not meant to be dismissive of current SSTEP scholarship; certainly teacher educators and their pre-service teachers have reaped great benefits from those types of studies, and they must remain as a key part of the growing field of SSTEP inquiry. While some examples of longitudinal research do exist in SSTEP (*cf.* Kosnik and Clift 2009), such studies are not common. Longitudinal, large-scale SSTEP (LLS-SSTEP) opens doors to other and different possibilities for PETE faculty. In our

18 years of conducting LLS-SSTEP at GSU, we have discovered that many promising possibilities can turn into valuable lessons learned through such a model; some of those possibilities were anticipated and served to drive our initial commitment, while other possibilities became apparent as the project continued over those years.

### ***Collaboration***

The simple reality of conducting SSTEP over long periods of time to collect and analyze large amounts of data and artifacts is that it quickly becomes much more work than any PETE professor can carry out alone. It also requires a greatly expanded base of expertise that is unlikely to be possessed by any single PETE professor. Collaboration is essential and raises certain possibilities for all involved. From our experience, PETEAP has fostered both internal and external collaboration. All of the GSU PETE faculty have willingly contributed individually to our institutional SSTEP efforts, and in doing that work we have found many mutual interests and learned much from and with each other. At other times our collective base of expertise was not adequate to allow us to pursue important SSTEP questions on our own. At various times we have sought the expertise of PETE faculty from other institutions to design and carry out parts of our SSTEP research agenda (Lund and Veal 2008a, b; McCullick 2008; McCullick et al. 2008; Mitchell 2000). From both the internal and external collaborations, we have formed a number of professional learning communities around various parts of the larger project.

### ***Deeper and Shared Understandings***

Not limited to a single research paradigm or to short-term analyses, LLS-SSTEP makes possible deeper understandings through the ongoing triangulation of data collected with different instrumentation and/or through different paradigms. The explanatory power of carefully triangulated evidence is much greater than that of discrete, disconnected evidence. Through collaboration and open discussion, these deeper understandings can be arrived at and shared by all faculty participants engaged in this process. At GSU we hold regular meetings to analyze and discuss our SSTEP data and artifacts so that all can contribute to a collective interpretation of what that evidence means for our program.

### ***Trend Analyses and Forward Planning***

The ongoing collection of data in LLS-SSTEP makes possible periodic analyses to detect trends in a PETE program over time. Some of these trends can be detected with casual observations by faculty, but many others lie hidden and cannot be seen

until data are carefully selected and analyzed in a time series. These analyses allow a PETE faculty to ‘use the past’ to ‘see the future’ and to either change or carry on accordingly. At GSU these trend analyses provide the basis for group discussions on topics such as program admission policies, course content and sequencing, field experience placements, teaching rubrics, and student teaching supervision assignments.

### ***Evidence-Based Decisions***

The collection of LLS-SSTEP data makes it possible for PETE faculty to collectively consider programmatic decisions from a solid base of evidence, especially when that evidence emanates from research quality data. This is in contrast to decisions that are made from limited, anecdotal, one-off studies derived from assessment quality data. We have found that our discussions around programming decisions are much more focused when they stem from trustworthy data, giving us greater confidence that the resulting decisions are as well-informed as possible.

### ***Examination of Programming Decisions***

Once a PETE faculty group has used existing data to make decisions and implement some type of change from that evidence, it becomes possible to analyze the efficacy of those decisions – if data on those decisions are collected after a change is enacted. This allows the group to determine if the change was in the desired direction and as effective as it was intended. It also provides some explanatory evidence for failed or less-robust planned changes in the program. As outlined by Metzler and Blankenship (1998, 2000) decisions by PETE faculty about their programs can be categorized within four levels: (1) Maintaining – not changing at all; (2) Adjusting – making single small changes; (3) Revising – making multiple large changes; and (4) Restructuring – changing the goals and/or major structural features of one or more parts. A study by Gurvitch and Metzler (2009) typifies how a restructuring decision led to a major positive outcome in our PETE program. In 2000 we changed the structure of the practicum experiences that PETE students had prior to entering formal student teaching placements. Specifically, we changed from practicum that were laboratory (on campus)-based to field-based, taught entirely in local schools with full classes of P-12 students and full-length class periods. From our longitudinal data we were able to measure the teaching efficacy (Gibson and Dembo 1984) of pre-2000 cohorts before and after formal student teaching and compare those data with cohorts from 2000 to 2005. What we learned is that the teaching efficacy of the laboratory-based cohorts was stronger upon entering student teaching, due to their success in the less complicated laboratory settings. However, their efficacy weakened considerably once the realities of student teaching were encountered.

Conversely, the efficacy of the field-based cohorts was lower going into student teaching due to the authenticity of those settings; their efficacy strengthened significantly during student teaching once they realized they had the skills and confidence to instruct well – having had more authentic pre-student teaching experiences.

## **Precautions for LLS-SSTEP**

While the GSU PETE faculty remain strong advocates for the conduct of LLS-SSTEP, we have learned much over these 18 years that can serve as sage advice for PETE groups who might consider using an ‘institutional SSTEP’ approach in their own programs. It should be noted that the following precautions are not meant to steer other PETE groups from doing LLS-SSTEP. Just the opposite: they are offered to provide some points to consider ahead of implementing such an approach locally.

### ***Do Not Start Without Full Disclosure and Commitment***

LLS-SSTEP is a massive undertaking; doing it well and sustaining it requires enormous amounts of time and effort by all involved. All PETE faculty/researchers need to have a firm understanding of the scope of the local project and express a commitment to it. We spent many months meeting to lay out the GSU PETEAP and formulated a draft of the project that was shared by all PETE faculty at the time. We all knew what we planned to do and each one of us was committed to the project before it started. As we recruited new faculty members for PETE later on, we were very clear to explain the project to those we interviewed, and sought an initial commitment to the project as a consideration in their hiring.

### ***Do Not Attempt This Alone***

LLS-SSTEP, by design, cannot be planned, implemented, and maintained by a single PETE professor/researcher, or even a small group of collaborators. The more minds involved at the planning stage, and the more hands available for the large amount of work needed during implementation, the better. Right from the start, this should be viewed as a shared, collaborative effort – and active contributors should be sought out at every stage. It would be advantageous to seek out external collaborators who have different expertise and who can lend new insights into the plan and eventual discussions and decisions.



### ***Do Not Run When You Start – Walk!***

Even if the initial LLS-SSTEP plan is ambitious, it will be best to implement the plan incrementally, demonstrating the ability to do each part well before pursuing additional parts of the plan. This strategy can avoid problems that eventually compound themselves as PETE faculty must divert attention from new research efforts while at the same time resolving issues embedded in ongoing work.

### ***Do Not Use Outdated Data Collection Technologies***

The GSU PETEAP started at a time of transition from data collected mostly by hand from hard-copy instrumentation to using electronic, on-line data collection technologies. Over the years we have shifted fully to collecting data with modern technology (e.g. SurveyMonkey); none of our quantitative data are presently collected from hand-copy versions of instruments. The use of electronic technology also prevents problems from having massive amounts of hard copy (paper) data to store, manage, secure, and analyze. This precaution might be obvious for today's technology-savvy PETE faculty members, but it needs to be expressed nonetheless.

### ***Do Not Re-invent the Data Instrumentation Wheel***

Regardless of what research questions are asked by PETE faculty today in SSTEP, it is very likely that valid and reliable instruments (quantitative) or accepted protocols (qualitative) already exist for that purpose; they can be used as presently designed, or with some simple modifications. It would behoove PETE faculty members to search those out first, rather than spending the considerable time and expertise needed to develop new instrumentation on their own.

### ***Do Not Be Parochial with LLS-SSTEP***

PETE faculty groups that pursue LLS-SSTEP will learn many things along the way, and have many experiences and insights to share with others. One of Galluzo and Craig's (1990) priorities for program assessment research is that it be used to inform other teacher educators, so that they may also benefit from what is learned by the 'local' group. Most often that benefit is derived from reading publications or attending presentations at conferences. If SSTEP data are of research quality (Metzler and Tjeerdsma 1998, 2000), there is a greater chance that those data will be disseminated in refereed scholarly outlets. Our commitment to collect research quality data has led

to many single data-based publications and two research monographs in the *Journal of Teaching in Physical Education* (Metzler and Tjeerdsma 1998; Gurvitch et al. 2008), thus sharing what we have learned with other PETEs around the world. As sharing is a key aspect of self-study research (Loughran 2005), dissemination becomes an important consideration in designing and conducting LLS-SSSTEP.

### ***Do Not Be Intimidated by External Demands for Data***

As mentioned earlier, at least in the United States, PETE faculty face onerous demands for providing program accountability data to a large number of external agencies. On the surface these demands may seem duplicative, but if the LLS-SSSTEP data-base is designed and managed well, duplication can be greatly reduced and even avoided. It has been our practice at GSU to design our data plan so it *first* produces the data we need for SSSTEP and simultaneously provides data needed for the many external reports we must generate annually.

### **Final Thoughts**

Since the PETEAP began at GSU in 1994, there have been over two dozen contributing researchers: GSU PETE faculty and graduate students, PETE faculty from other universities, and P-12 teachers – many of whom graduated from our pre-service teacher education program. It has truly been a collaborative effort, in the very best sense of that term. In different combinations over the years, we have maintained an ongoing professional learning community, centered in the project and dedicated to gathering usable evidence to make informed decisions about our pre-service PETE program. And, we have been able to conduct that inquiry in a manner that has allowed us to disseminate it to other PETE professionals in the US and abroad. It has never been easy, and it has sometimes not been perfect, but the work we have done to achieve a longitudinal, large-scale institutional self-study of teacher education practice in PETE has been well worth the toil. In the end, we have all learned more about PETE through this collaborative effort than we could have possibly learned through individual inquiry.

### **References**

- Erickson, L. B., Pinnegar, S., & Young, J. R. (2012). A programmatic self-study of practice: Exploring teacher educator knowledge. In J. R. Young, L. B. Erickson, & S. Pinnegar (Eds.), *The ninth international conference of self-study of teacher education practices* (pp. 98–101). Provo: Brigham Young University Press.
- Galluzo, G. R., & Craig, J. R. (1990). Evaluation of preservice teacher education programs. In W. R. Houston, M. Haberman, & J. Sikula (Eds.), *Handbook of research on teacher education* (pp. 599–616). New York: Macmillan.

- Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Education Psychology, 76*, 569–582.
- Gurvitch, R., & Blankenship, B. J. (2008). Chapter 6: Implementation of model-based instruction – the induction years. In R. Gurvitch, M. Metzler, & J. Lund (Eds.), *Model based instruction for physical education: The adoption of innovation. The Journal of Teaching in Physical Education, 27*, 529–548.
- Gurvitch, R., & Metzler, M. (2009). Teaching efficacy in preservice physical education teachers. *Teaching and Teacher Education, 25*, 437–443.
- Gurvitch, R., Metzler, M., & Lund, J. (Eds.). (2008). Model based instruction for physical education: The adoption of innovation. *The Journal of Teaching in Physical Education, 27*, 447–589.
- Hamilton, M. L. (2002). Change, social justice, and reliability: Reflections of a secret (change) agent. In J. Loughran & T. Russell (Eds.), *Improving teacher education practices through self-study* (pp. 176–189). London: RoutledgeFalmer.
- Kosnik, C., & Clift, R. T. (Eds.). (2009). Longitudinal self-studies of teacher education practices. *Studying Teacher Education, 5*, 103–199.
- Loughran, J. (2005). Researching teaching about teaching: Self-study of teacher education practices. *Studying Teacher Education, 1*(1), 5–16.
- Loughran, J. (2010). Searching for meaning in structuring preservice teacher education. In L. Erickson, S. Pinnegar, & J. Young (Eds.), *Navigating the public and private: Negotiating the diverse landscape of teacher education* (pp. 133–136). Proceedings of the eighth international conference on the self-study of teacher education practices, Herstmonceux Castle, East Sussex, England. Provo, UT: Brigham Young University.
- Lund, J. L., & Veal, M. L. (2008a). Chapter 7: Influences on cooperating teachers' adoption of model-based instruction. In R. Gurvitch, M. Metzler, & J. Lund (Eds.), *Model based instruction for physical education: The adoption of innovation. The Journal of Teaching in Physical Education, 27*, 549–570.
- Lund, J. L., & Veal, M. L. (2008b). Measuring pupil learning: How do student teachers assess within instructional models? In R. Gurvitch, M. Metzler, & J. Lund (Eds.), *Model based instruction for physical education: The adoption of innovation. The Journal of Teaching in Physical Education, 27*, 487–511.
- McCullick, B. (2008). Assessing a PETE program through the eyes of cooperating teachers. In M. Metzler & B. Tjeerdsma (Eds.), *The Georgia State University Physical Education Teacher Education Program Assessment Project. Journal of Teaching in Physical Education, 19*, 508–521.
- McCullick, B., Metzler, M., Cicek, S., Jackson, J., & Vickers, B. (2008). Kids say the darndest things: PETE program assessment through the eyes of children. *Journal of Teaching in Physical Education, 27*, 4–20.
- Metzler, M., & Tjeerdsma, B. (Eds.). (2000). The Georgia State University Physical Education Teacher Education Program Assessment Project. *Journal of Teaching in Physical Education, 19*, 399–555.
- Metzler, M., & Tjeerdsma, B. (1998). PETE program assessment within a development, research, and improvement framework. *Journal of Teaching in Physical Education, 17*, 468–492.
- Mitchell, M. (2000). An approach to program assessment: Locating indicators of a coherent program. In M. Metzler & B. Tjeerdsma (Eds.), *The Georgia State University Physical Education Teacher Education Program Assessment Project. Journal of Teaching in Physical Education, 19*, 522–537.
- National Association for Sport and Physical Education. (1995). *National standards for beginning physical education teachers*. Reston: Author.
- Zeichner, K. M. (2007). Accumulating knowledge across self-studies in teacher education. *Journal of Teacher Education, 58*(1), 36–46.