

**Valuing Teachers' Evaluative Thinking:
The Role of Teacher Knowledge and Practice in Formative Assessment**

*Leigh M. Tolley, Ph.D.
University of Louisiana at Lafayette*

Abstract

On a daily basis, PreK-12 teachers make countless decisions about how to best meet their students' needs through assessment and instruction. However, their ability to justify these decisions is not frequently used as a resource for accountability measures or the evaluation of educational programs. The research presented here, which draws from a sequential explanatory mixed methods study of seven secondary English/language arts teachers' formative assessment practices, delves into how teachers' evaluative thinking occurs in the classroom. Data were collected using the experience sampling method with a series of weekly self-report checklists, as well as semi-structured interviews. Through examining emergent themes from this study and investigating the role of evaluation in education, implications are made for the importance of valuing this process and making this implicit teacher knowledge more explicit.

Keywords: educational evaluation, teacher knowledge, formative assessment

Introduction

In teacher preparation programs, preservice teachers rely upon their cooperating teachers to provide guidance on management, lesson planning, instruction, and assessment. Conversations that allow for open discussion about classroom practices are essential to new teachers' development. Likewise, experienced teachers, whether acting formally as instructional coaches or more informally as peer feedback providers, often participate in reciprocal learning that benefits all involved in their professional growth. While conducting a study with a focus on secondary English/language arts (ELA) teachers' formative assessment practices (Tolley, 2016), I observed that teachers' evaluative thinking emerged as an unanticipated theme. This article draws from that prior research and builds upon the idea of valuing teachers' capacity and skill for evaluative thinking, both through the lens of examining formative assessment in a secondary ELA context and in the field of education as a whole.

Theoretical Framework

The theoretical framework that guided this research was grounded in the fields of evaluation and education, specifically drawing from elements of program evaluation, conceptions of teacher knowledge and practice, and formative assessment in the classroom. Each of these concepts can be fairly broad in scope, especially the term "formative assessment" and how it may be operationalized in practice. By bringing the three areas together to create a unified and specific perspective through which this research could be studied, there was a targeted focus on understanding teachers' evaluative thinking in a secondary ELA context. A brief review of relevant literature from each of these areas is presented below.

Understanding Evaluation

Evaluation is primarily defined as "the process of determining the merit, worth, or value of something, or the product of that process" (Scriven, 1991, p. 139). As a field, evaluation applies to all contexts and subjects,

and has therefore been noted for its “multidisciplinarity.” Its focus on the investigation of myriad aspects of what is being evaluated—known as the *evaluand*—differs from the directed hypothesis-testing nature of social science research (Scriven, 1991). One form of evaluation, *program evaluation*, is frequently used in education through the collection of data to better understand a program in which teachers, students, administrators, and other stakeholders are involved, as well as products and processes involved in the program, and how well it serves its intended audience. The terms “formative evaluation” and “summative evaluation” were introduced by Scriven (1967) to describe two different evaluation types. The former is used to describe an evaluation that occurs during the development or improvement of a program, product, or person, often conducted more than once, with the intent to improve. The latter refers to an evaluation conducted after completion of the program for the benefit of an external audience or decision-maker (Scriven, 1991). Both of these terms have been adapted into education through the terms *formative assessment* and *summative assessment*. Formative assessments are meant to guide and improve student learning as well as teachers’ instructional practice, while summative assessments are used for reporting purposes (Black, 2013; Bonner, 2013).

The importance of context. In the evaluation of educational programs, the context in which a program is being implemented greatly impacts its outcomes (Tolley, 2011). For example, there may be contextual variance at the administrative level, which may include perceived pressure to meet external accountability measures (e.g., Black & Wiliam, 2005; Nelson & Eddy, 2008); at the school level, in terms of professional development and collaboration time allotted to staff (e.g., Frey & Fisher, 2009; Nelson & Eddy, 2008; Wiliam, 2006); at the classroom level, which may include pedagogical differences and teacher fidelity of implementation to a program (e.g., Crossouard, 2011; Yin et al., 2008); and at the student level, such as demographics and individual achievement (e.g., Crossouard, 2011; Frey & Fisher, 2009; Nelson & Eddy, 2008; Yin et al., 2008). Each of these differences must be accounted for when evaluating a program’s merit, especially those that occur within a classroom.

Evaluative thinking. In their classrooms, teachers analyze and use the results of their informal and formal assessment data to guide instructional decision making, and these choices are a form of evaluative thinking (Nelson & Eddy, 2008). Evaluative thinking in the past has been defined as “essentially critical thinking applied to contexts of evaluation” (Buckley, 2015, as cited in Patton, 2018, p. 12), but it has more recently been argued that critical thinking is one of many components of this concept. These diverse aspects include the ongoing evaluation associated with Freirean pedagogy; the strengthening of democracy as education, a concept supported by Dewey and other thought leaders; the skills and abilities associated with higher-order thinking in Bloom’s *Taxonomy of Educational Objectives*; and its use in the field of evaluation as a whole through logic, argument, and rigorous thinking (Patton, 2018). The principles of evaluative thinking, which include the need to be clear, intentional, accountable, specific, and systematic; to focus and prioritize, and make assumptions and criteria and standards for judgment explicit; use data-based statements of fact and limit generalizations and causal explanations to what data support in order to draw appropriate conclusions; and the need for cultural sensitivity and cultural competence (Patton, 2018) can all quite easily be connected to what are considered to be effective best practices in teaching.

Conceptions of Teacher Knowledge and Practice

When considering teachers’ classroom practices, it is imperative to take their knowledge into account. Teachers’ decision making is often an implicit process, and capturing and making this knowledge explicit would allow teachers to learn from their peers and share their ideas with others, including preservice teachers that are being mentored by them.

What does teacher knowledge entail? According to Shulman (1987), teacher knowledge, at a minimum, includes: (a) content knowledge; (b) general pedagogical knowledge, including classroom management and

organization; (c) curriculum knowledge; (d) pedagogical content knowledge; (e) knowledge of learners and their characteristics; (f) knowledge of educational contexts, from classroom through culture; and (g) knowledge of educational purposes and values and their philosophical and historical grounds. Of particular note is pedagogical content knowledge, “that special amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding” (p. 8). Shulman asserted that teachers’ knowledge is developed from at least four major sources, which he identified as scholarship in content areas, the materials and structure of teaching, ongoing research in the field, and the wisdom that comes from conducting the practice itself.

How do we learn from this knowledge? Expert teachers, or those with strong practical knowledge, have a certain set of features that is similar to experts in other fields. In part, these features include excelling within particular contexts, being more opportunistic and flexible in their teaching than novices, and perceiving more meaningful patterns in the domain in which they are experienced (Berliner, 2001). Although the time for the development of expertise may vary, “a reasonable estimate for expertise to develop in teaching, if it ever does, appears to be 5 or more years” (p. 479). When studying teacher practice, written questionnaires may reveal relevant data, but more in-depth information can be gleaned from teachers in narrative form, such as through interviews where teachers reflect on their practice (e.g., Clemente & Ramírez, 2008; McMillan, 2003) to learn about their decision-making processes.

Formative Assessment in the Classroom

Formative assessment (FA) is a term “encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged” (Black & Wiliam, 1998, pp. 7-8). Four major components of FA have emerged from studies of the effectiveness of this concept. The first component is questioning students appropriately in ways that engage their thinking strategies, allow for response time, and act as a means of gathering evidence to inform instruction (e.g., Black, Harrison, Lee, Marshall, & Wiliam, 2003; Shermis & Di Vesta, 2011). The second component, feedback, whether from teachers to students, students to peers, or students to themselves in the form of self-regulation, is another important piece of effective FA (e.g., Black et al., 2003; Heritage, 2010; McManus, 2008; Popham, 2008; Sadler, 1989; Shermis & Di Vesta, 2011). The third and fourth components both address assessment materials: the construction of effective assessments and the analysis and use of assessment data to drive instructional decision making (e.g., Birenbaum, Kimron, Shilton, & Shahaf-Barzilay, 2009; Griffin, 2009; Nelson & Eddy, 2008; Phelan, Choi, Vendlinski, Baker, & Herman, 2011; Yin et al., 2008).

In an attempt to operationalize the complex construct of FA within education, I conducted a literature review of 25 different studies on professional development for teachers in grades K-12. Although definitions and implementations of FA varied, common elements included: (a) teachers setting and making explicit clear learning goals, (b) the use of questioning and feedback from teachers and students, (c) data collection, analysis, and use; (d) teacher revision/modification of instruction, (e) the development of student academic autonomy, and (f) evaluation to determine if a gap between the learning goals and current student performance exists (Tolley, 2012). Like evaluation practice, FA implementation in the classroom varies based on the context in which it takes place.

For FA in a secondary English/language arts (ELA) context, there are anecdotal reports of practices that teachers use with their students, but a lack of empirical data about FA implementation. In part, this is due to assessment within ELA being more individualized, organic in nature (Marshall, 2007), and in-the-moment, without as many pre-planned responses and feedback opportunities when compared to more structured, sequential content areas such as math and science (McMillan, 2010). A stronger base of empirical data—and the reasoning behind them—would help future educational research and evaluation endeavors to be grounded in what actually occurs at the classroom level.

Bringing together Program Evaluation, Teacher Knowledge, and Formative Assessment

By uniting the three concepts of program evaluation, teacher knowledge, and formative assessment for this study, a unique lens was used to view this research in a way that was context- and content-specific. This particular approach was focused on what secondary ELA teachers use in practice in constructs that are otherwise fairly nebulous and challenging to clearly define, with an aim of contributing to empirical research in these areas.

Methods

The study from which this research emerged used a sequential explanatory mixed methods design (Creswell & Plano Clark, 2011; Ivankova, Creswell, & Stick, 2006), with the qualitative phase built upon the quantitative phase to help to explain the findings (Ivankova et al., 2006). The larger study addressed the following four research questions:

1. What assessment practices do secondary ELA teachers use?
2. How do secondary ELA teachers determine what assessment practices to use and when to use them?
3. What are secondary ELA teachers' perceptions of the effectiveness of the assessment practices that they use?
4. How do secondary ELA teachers determine the effectiveness of their assessment practices?

For this study into teachers' evaluative thinking, an additional research question was addressed by examining the findings from questions 2 and 4, above:

5. How does secondary ELA teachers' discussion of their assessment practices show evidence of evaluative thinking?

Participants

Study participants were ELA Instructors in a long-standing concurrent enrollment program that partners secondary schools with a private university in the northeastern United States. The Instructors are teachers in the schools that are trained by the program, and high school students can earn college credits through successful completion of program coursework. To be an Instructor in this program, teachers must have applied for and hold specific credentials, including at least a master's degree, and participate in both initial and ongoing training and professional development.

Instrumentation

There were three instruments that were all created for the purposes of the initial research: (a) a background questionnaire, (b) the English/Language Arts Teachers' Experiences (ELATE) checklist, and (c) a semi-structured interview protocol. More information about each of these instruments is presented below.

Background questionnaire. The eight-item background questionnaire asked for participants' gender and years of teaching experience as of Spring 2015, both generally and in the concurrent enrollment program, and their teaching load at the time of this study (2014-2015) in terms of students and sections/courses taught.

English/Language Arts Teachers' Experiences (ELATE) checklist. The ELATE checklist contained 25 items (see Appendix A)—a list of assessment practices related to questioning, discussion, and feedback—as well as space for Instructors to indicate other assessment practices that they used which were not included in the list. Participants were asked to complete one checklist per week, documenting their assessment practices from Friday through Thursday, for four weeks. These days were chosen instead of Monday through Friday so as to not overburden teachers at the end of the work week. This administration of weekly checklists was an experience sampling method (e.g., Hektner, Schmidt, & Csikszentmihalyi, 2007; Zirkel, Garcia, & Murphy, 2015), which allows for an understanding of individuals' experiences in context as they unfold.

Instructors were asked to indicate how many times over the past week they used each practice with their concurrent enrollment program students. For the practices that they used, the Instructors were asked to rate their perceived effectiveness of that practice to improve student learning with their lower-performing, average-performing, and higher-performing program students using a 6-point Likert-type scale (0 = *extremely ineffective*, 1 = *very ineffective*, 2 = *moderately ineffective*, 3 = *moderately effective*, 4 = *very effective*, and 5 = *extremely effective*). The checklist items were based on the previously-mentioned literature review of studies of professional development on FA (Tolley, 2012); Danielson's *Framework for Teaching* (2007, 2014), particularly the "using questioning and discussion techniques" and "using assessment in instruction" components of the instruction domain; the questioning techniques section of the SAIP Writing Assessment III Teacher Questionnaire (Hunter, Mayenga, & Gambell, 2006); and an earlier fieldwork study that had been conducted with this concurrent enrollment program.

Semi-structured interview protocol. An individual semi-structured interview (e.g., Rubin & Rubin, 2005) was conducted with participants after their completion of the weekly checklists. The interview protocol was finalized after the background questionnaire and ELATE checklists data (i.e., the quantitative data) were analyzed, and was created to include individualized questions specific to each Instructor (i.e., the qualitative stage of the study). The participant-specific questions added to the general protocol were intended to check the validity of ELATE checklist responses and request clarification of why certain assessment practices were used (or not) or deemed effective (or ineffective). In addition, questions were asked of participants to help them reflect on their perceptions of assessment in their teaching after having completed the series of ELATE checklists and thinking about their own practices.

Data Analysis

Data from the quantitative phase of this study (background questionnaires and ELATE checklists) were entered into Microsoft Excel and IBM SPSS Statistics, v. 22. Descriptive statistics (e.g., frequencies, means, and standard deviations) were calculated, as were chi-square tests, adjusted one-way repeated measures analysis of variance (ANOVA) with post hoc analysis, and Pearson correlations for the ELATE checklist items (Leech, Barrett, & Morgan, 2011; Morgan, Leech, Gloeckner, & Barrett, 2013; Sprinthall, 2007). Participants' responses to open-ended items of this phase were reviewed and coded for emergent themes.

For the qualitative phase of the study, all interviews were audio recorded with the participants' permission, and were transcribed verbatim into Microsoft Word format with the aid of Dragon NaturallySpeaking, v. 13. Once transcribed, the qualitative data were reviewed and coded for emergent themes (Creswell, 2008; Stake, 2010) using constant comparative analysis (Onwuegbuzie, Dickinson, Leech, & Zoran, 2009; Strauss & Corbin, 1998).

Findings

Participants and Their Contexts

The seven Instructors in this study were all teaching the same semester-long college-level writing course in Spring 2015, and their characteristics are presented in Table 1. Of the seven teacher participants, six were female and one was male. Their total years of teaching experience ranged from 10 to 27 years ($M = 18.14$, $SD = 6.23$), with between 3 and 15 years ($M = 8.14$, $SD = 3.77$) of teaching in the concurrent enrollment program.

Although all participants were teachers in the same state in the northeastern United States, the context of their different schools varied. Table 2 presents the characteristics of the teachers' schools at the time of the study.

Table 1
Characteristics of Teachers, Spring 2015

Teacher	Gender	Total Years Teaching	Years Teaching in Program
1. Mia	Female	17	7
2. Helen	Female	27	15
3. Allison	Female	25	10
4. Charlotte	Female	20	6
5. Rachel	Female	15	3
6. Krystal	Female	13	9
7. Tony	Male	10	7

Table 2
Characteristics of Teachers' Schools

Teacher	School Locale	Grade Levels of School	Total Students in School	Enrollment by Gender		Students Eligible for Free/Reduced-price Lunch (% of Total)
				Male	Female	
1. Mia	Suburb: Large	9-12	1,659	858	801	124 (7.47)
2. Helen	Suburb: Large	9-12	2,186	1,149	1,037	54 (2.47)
3. Allison	Rural: Fringe	9-12	1,526	763	763	125 (8.19)
4. Charlotte	Suburb: Large	9-12	1,029	542	487	152 (14.77)
5. Rachel	Rural: Fringe	7-12	716	377	339	272 (37.99)
6. Krystal	Suburb: Large	9-12	2,263	1,163	1,100	575 ^a (25.41)
7. Tony	Suburb: Large	9-12	1,600	788	812	99 (6.19)

Note. Data from CCD Public school data, 2013-2014 school year (National Center for Education Statistics, 2016).

^a Data are from 2012-2013 school year, as 2013-2014 data did not meet NCES quality standards.

All of the students who were in the concurrent enrollment program ELA course taught by the study participants were in 12th grade. The prerequisites for the students to be in this course varied by school. For example, Mia, Helen, and Allison noted that there were prerequisite courses, and that the students that enrolled in the concurrent enrollment program ELA course typically took honors or Advanced Placement courses. Charlotte, however, stated that her school had an open enrollment policy, so any student that was interested in taking the concurrent enrollment program ELA course was permitted to do so.

Of the seven participants, all of them completed the series of ELATE checklists. However, Tony elected not to participate in the semi-structured interview.

Factors Impacting Teachers' Selection of Assessment Practices

The data from the ELATE checklists and semi-structured interviews indicated that there were four major emergent themes that affected teachers' selection of assessment practices to use with their students (Research Question 2). These were (a) the purpose of assessment, (b) the intertwining roles of assessment and instruction, (c) the teachers' own knowledge and experience, and (d) the classroom context as factors that influenced their decision making.

The purpose of the assessment. Teachers considered the end goals and objectives for their use of assessment practices. In interviews, several of the participants described the need to meet their students' needs, especially for improving student learning, or how they gauged student understanding within the classroom. Meeting short-term teaching goals for students and fairness in student evaluations were other factors that emerged.

Helen, one of the participants, defined assessment as “an honest attempt to evaluate how much learning has gone on.” She also remarked that assessment “should never be hard and fast, it should never be definitive for the sake of being definitive” (interview dated June 26, 2015).

Assessment and instruction. Another major theme that emerged in the discussion of participants' selection of their assessment practices was the connection of the practice to instruction. Teachers reported using the information that they learned from their assessments to guide what they did with their students. In addition, they made distinctions between informal classroom assessments and higher-stakes assessments, such as those required by the state.

For example, Allison shared that “I use assessment in different ways. One is *diagnostic*, and, um, the other is *evaluative*, I guess, to determine how — how did they perform in this particular class” (interview dated July 1, 2015). She further explained how she used assessment in her instruction:

a lot of assessment, you *don't* give a grade on, so, it may be — like, in discussion in class, you will realize that they haven't got a concept that you want them to get, or they didn't really understand a text that you read together. So, that maybe, is diagnostic [chuckles] in terms of even *my* teaching practices, so maybe I have to go back and spend more time making sure they get it. (interview dated July 1, 2015)

Teachers' own knowledge and experience. Teachers drew upon their knowledge as educators to make assessment decisions. At the time that this study was conducted, all of the teacher participants had been educators for at least 10 years and had been concurrent enrollment program Instructors for at least three years. Of these participants, Helen had the most teaching experience, having taught for 27 years overall, including 15 years as a program Instructor. In our interview, she discussed the kind of thinking about her students that has emerged from her experiences:

I think all teachers probably, even before the first paper comes in you, you — you're figuring, you know, based on classroom performance and discussion and all that, you're kind of getting an idea of who's going to, you know — who's, who's going to be the better writers, who's going to have more to say, and all of that sort of thing. Um, I think...I think you just do a lot of, um, evaluating — I think you're evaluating all of the time. (interview dated June 26, 2015)

The other teachers expressed similar sentiments, especially when discussing student performance levels in their concurrent enrollment program ELA classes. Charlotte explained how her reflection on her practice and her teaching experience impacted how she thought about assessment:

There are some things in the moment; after teaching long enough, you can usually figure out, ‘If that’s not working, what else can I do right now,’ um, ‘to try to achieve that?’ And again, sometimes it’s assessing how well my goal is met versus how well they are understanding or grasping something. There’s a — I think a, a lot of multitasking happening with assessment: it’s both, um, self-assessment, but also assessment of students. (interview dated June 29, 2015)

In addition, several of the interviewed teachers remarked that participation in the study made them more aware of their reflective practices, and more conscious of their decision-making processes. For example, when talking about her understanding of “formative assessment,” Mia expressed how integral the study of and reflection on assessment was to her identity as a teacher, saying, “it’s important to me [...] I think it’s just the nature of what I do” (interview dated July 17, 2015). Rachel remarked that as she was completing the checklists each week, she thought to herself, “OK, now, I’m cognizant of this, so I’m now sort of watching [...] how many times [I’m] doing this, and what does this mean” (interview dated June 25, 2015).

The classroom context. During our interview, Helen said the following about the complex environment of education: “assessment never stands...you know, it’s not like a scientific principle. It, it never stands outside of context” (interview dated June 26, 2015). The context of a classroom and its impact on how study participants chose which assessment practices to use was a fourth major emergent theme, and included the following minor themes: (a) how teachers established expectations for their students, (b) the rapport and collegiality that existed between individuals, (c) the interactions that occurred between students and their peers, and (d) the different types of student learners that were present in the classroom.

Here, too, knowledge of students and their abilities and preferences played a role in the teachers’ assessment selections. Krystal discussed how she adjusted her assessment practices to meet her individual students’ needs, and stated that for some of her learners, “I know they get it, and I know they can demonstrate things better orally” than in writing; at the same time, she had other students who “don’t like to give me things in writing” and who might be better served by “at least check-in in terms of discussion.” She explained that she learned the best methods for assessing students’ learning progress based on her experiences with each person, saying, “I usually kind of start with the same approach at assessment, and then as I go throughout the courses, I get to know them better, and then I can kind of make adjustments as I need to” (interview dated July 9, 2015).

Factors Impacting Teachers’ Perceptions of Effectiveness of Assessment Practices

There were four major themes that emerged from the data in the larger study about factors affecting teachers’ perceptions of the effectiveness of assessment practices (Research Question 4). Three of these were students’ understanding of content, student performance levels, and the potential benefits of the assessment practice to the students. Most relevant to the current research, however, is a fourth major emergent theme: teachers’ knowledge of students. This particular theme is explained in more detail below in terms of (a) familiarity with individual students, and (b) ability to address students’ needs through assessment and instruction.

Familiarity with individual students. All of the interviewed study participants discussed their knowledge of their concurrent enrollment program students and their backgrounds; this knowledge tended to be even deeper if the teacher had previously taught a particular student. For instance, Charlotte had many of her program students in her ELA courses the year before. She said of this, “I came in [to the program’s ELA course], you know, with pre-existing relationships and already knowing something about a lot of them, not all of them, but a lot of them and their writing” (interview dated July 9, 2015).

Another example is Krystal, who when asked how she makes decisions about which assessment practices to use for her students, replied, “I kind of get a sense [of what each student’s needs are] — and I’m, again, I’m reasonably lucky in [the program] that I’ve had many of them before” (interview dated July 9, 2015). Her familiarity with these students also permitted her to tailor her assessment practices to best suit them as learners.

Ability to address students’ needs through assessment and instruction. Interviewed participants also indicated that they used their knowledge of individual students in both assessment and instruction. Helen remarked that “assessment, I think in the beginning, is — as much your, your gut instinct as to who you’re talking to. You know, who your learners are [...] So I think assessment is — it’s never one-size-fits-all” (interview dated June 26, 2015).

In a similar vein, Krystal commented that “assessment’s main function is, you know, kind of in that individual level of each kid, how well they’re understanding, but more holistically.” She further explained that “If I have a lot of students in a given class who aren’t getting a particular concept, well then, that’s telling me that perhaps that’s a concept I need to revisit, that something went wrong in my delivery, um, or in my setup, or in my scaffolding, you know, whatever it is, and that’s something I need to revisit. (interview dated July 9, 2015)

Evidence of Evaluative Thinking

In addition to the findings above that demonstrated evaluative thinking in how study participants described their selection of assessments and the determination of their efficacy (Research Questions 2 and 4, respectively), more evidence of teachers’ evaluative thinking (Research Question 5) emerged from the interviews through two other themes: (a) evaluative thinking that informed instruction, and (b) evaluative thinking internally and externally.

Evaluative thinking that informed instruction. Rachel described her approach to assessment as changing as she became more experienced as a teacher:

early on, I think really for me [...] assessments [...] were more shaped for data collection, [and now] I, you know, *think* about them, I see them as a *resource*. I see them as a *guide*, I see them as very *useful* tools — not that they weren’t useful before, but it’s in *how* I look at what I’m getting from kids that causes me to change what I do. (interview dated June 25, 2015)

Rachel went on to add:

I think in this whole data-driven society, it’s more — [...] we’re looking at the numbers, we’re looking at the numbers, but we kind of need to step away from that, and just say, ‘OK, now — *what’s* relevant, and how — *how* may I use this to assist students? How may I use this to help *me* as instructor?’ [...] I had to step away from ‘This is the information that I’m getting,’ to, it’s not just about that; it’s about ‘How does the information that I’m getting *translate* to what maybe — you know, what I need to *say* again, what I need to *do* again, what, you know, *evidence*, do I talk to my colleagues about?’ (interview dated June 25, 2015)

Mia also described how she used evaluative thinking in her own teaching of her students, stating “what I’m basically looking at is what *have* they learned? What have — what have they absorbed, and now, how can I *add* to that next time?” (interview dated July 17, 2015).

Evaluative thinking internally and externally. Charlotte gave examples of some of the questions that she asked herself as part of the assessment process:

There are some things in the moment; after teaching long enough, you can usually figure out, ‘If that’s not working, what else can I do right now to try to achieve that?’ And again, sometimes it’s assessing how well my goal is *met* versus how well [the students] are understanding or grasping something. There’s a —

I think a, a *lot* of multitasking happening with assessment: it's both, um, self-assessment, but also assessment of students. (interview dated June 29, 2015)

Krystal described her internal evaluative thinking becoming explicit, or external, when working with a student teacher from the university supporting the concurrent enrollment program:

when you have a student teacher, you are [says next few words slowly for emphasis] thinking out loud, constantly. And I think, in demonstrating the 'how' and the 'why I do what I do,' it really has made me more cognizant of my practice overall, whether I have a student teacher there or not: it makes me think about it more. [...] When you have to think out loud, you're — I guess you knew why you were doing a thing, but explaining it to somebody else...? That's a whole different ballgame. (interview dated July 9, 2015)

Discussion

Making Teacher Knowledge Explicit

The teachers in this study, through discussing their FA practices within a secondary ELA context, were able to make their implicit knowledge explicit. Their years of experience as educators, as well as involvement in a program that included training and ongoing professional development, allowed them to develop methods that were well-suited to addressing their students' needs. This focus on students and ways to use assessment to improve their learning relates to Heritage (2013) and her explanation of how classroom teachers develop and use a learning progression for their students to help them learn content-specific knowledge and skills. She wrote that this progression

is based on [the educator's] experience of teaching children. Their sources for developing the progression are curricula, their views of what is best taught when, and their knowledge of children's learning. In this context, validation involves working together, testing each other's hypotheses against their professional knowledge, making refinements accordingly, trying out the progression to see if their model actually predicts what happens in terms of student learning, and then making further refinements from this experience. A by-product of teacher-developed progressions is an associated deepening of teacher knowledge about learning in a domain, which can have considerable payoff for evidence gathering and use. (p. 189)

Teachers' Evaluative Thinking and Formative Assessment

Helen's comment that "I think you're evaluating all of the time" (interview dated June 26, 2015) seemed to be representative of the interviews with the other participants in this study and their thoughts about assessment. The continuous, reflective, and recursive processes that teachers experience in the classroom through FA practices involve many different of-the-moment decisions, all with an aim to increase student knowledge. As a result, teachers are likely more connected to the field of evaluation than is currently acknowledged. FA involves teachers using evaluative skills and strategies to improve student learning through timely feedback, it is akin to program evaluators offering recommendations for improvement (Ayala & Brandon, 2008). Continued study of FA practices, including teachers' evaluative thinking, questioning, discussion, and feedback, and how they inform teachers' instructional decision making—in ELA and other content areas—can ultimately help to inform educational program evaluation, especially for formative evaluation purposes.

Study Limitations

The primary limitations for this study were the small sample size and the use of self-reported data from teachers. Although there were only seven participants, six of whom participated in the interviews, the larger study was focused on researching a particular population as a means of determining the feasibility of learning about

secondary ELA teachers' assessment practices from expert instructors, rather than the generalizability of the results. The data from this study were self-reported by the participants, which raises concerns about the reliability and validity of the findings. The use of experience sampling methods (Hektner et al., 2007; Zirkel et al., 2015) alleviated some of these concerns, as this approach asks participants about something that just happened and serves to capture events and actions that would be challenging to capture otherwise. In addition, teacher self-reports that are focused (e.g., specified subject matter, class group, and time frame for retrospective reporting) can be used to gather reliable data about their practices (Koziol & Burns, 1986).

Conclusion

Evaluators and researchers that collaborate with teachers to evaluate, study, or even better understand programs can be more proactive in discussing evaluative thinking and FA practices under study, and benefit from the knowledge that teachers already have. Teachers, especially those that are experienced and continually involved in relevant professional opportunities to share and reflect on their practices, are well-suited to explain what may and may not work in a given instructional context. These teachers can more clearly explain the reasoning for selecting assessment practices based on their purpose and connection to instruction that would be contextually appropriate for their students. In addition, experienced teachers' use of their knowledge of their students to judge the effectiveness of assessment practices and their internal and external evaluative thinking processes related to both assessment and instruction would be beneficial to share and discuss with other stakeholders to best suit students' needs. By bringing together teachers, evaluators, and researchers to develop common means of discussing what goes on in the classroom, there will be improved understanding of pedagogy, teacher decision making and the reasoning behind those choices, and fidelity of implementation of programs in the classroom, which all will ultimately lead to improved student learning.

References

- Ayala, C. C., & Brandon, P. R. (2008). Building evaluation recommendations for improvement: Insights from student formative assessments. In N. L. Smith & P. R. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 159-176). New York, NY: The Guilford Press.
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35(5), 463-482.
- Birenbaum, M., Kimron, H., Shilton, H., & Shahaf-Barzilay, R. (2009). Cycles of inquiry: Formative assessment in service of learning in classrooms and in school-based professional communities. *Studies in Educational Evaluation*, 35(4), 130-149.
- Black, P. (2013). Formative and summative aspects of assessment: Theoretical and research foundations in the context of pedagogy. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 167-178). Los Angeles, CA: SAGE Publications.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Berkshire, England: Open University Press.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.
- Black, P., & Wiliam, D. (2005). Lessons from around the world: How policies, politics and cultures constrain and afford assessment practices. *Curriculum Journal*, 16(2), 249-261.
- Bonner, S. M. (2013). Validity in classroom assessment: Purposes, properties, and principles. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 87-106). Los Angeles, CA: SAGE Publications.
- Clemente, M., & Ramírez, E. (2008). How teachers express their knowledge through narrative. *Teaching and Teacher Education*, 24(5), 1244-1258.

- Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, NJ: Pearson.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: SAGE Publications.
- Crossouard, B. (2011). Using formative assessment to support complex learning in conditions of social adversity. *Assessment in Education: Principles, Policy & Practice*, 18(1), 59-72.
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- Danielson, C. (2014). *The Framework for Teaching evaluation instrument: 2013 edition*. Princeton, NJ: The Danielson Group, LLC. Retrieved from <http://danielsongroup.org/download/?download=448>
- Frey, N., & Fisher, D. (2009). Using common formative assessments as a source of professional development in an urban American elementary school. *Teaching and Teacher Education*, 25(5), 674-680.
- Griffin, P. (2009). Teachers' use of assessment data. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational assessment in the 21st century: Connecting theory and practice* (pp. 183-208). London, England: Springer.
- Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks, CA: SAGE Publications.
- Heritage, M. (2010). *Formative assessment and next-generation assessment systems: Are we losing an opportunity?* Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Retrieved from http://www.edweek.org/media/formative_assessment_next_generation_heritage.pdf
- Heritage, M. (2013). Gathering evidence of student understanding. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 179-195). Los Angeles, CA: SAGE.
- Hunter, D., Mayenga, C., & Gambell, T. (2006). Classroom assessment tools and uses: Canadian English teachers' practices for writing. *Assessing Writing*, 11(1), 42-65.
- Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, 18(3), 3-20.
- Koziol, S. M., Jr., & Burns, P. (1986). Teachers' accuracy in self-reporting about instructional practices using a focused self-report inventory. *The Journal of Educational Research*, 79(4), 205-209.
- Leech, N. L., Barrett, K. C., & Morgan, G. A. (2011). *IBM SPSS for intermediate statistics: Use and interpretation* (4th ed.). New York, NY: Routledge.
- Marshall, B. (2007). Formative classroom assessment in English, the humanities, and social sciences. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 136-152). New York, NY: Teachers College Press.
- McManus, S. (2008). *Attributes of effective formative assessment*. Washington, D.C.: Council of Chief State School Officers. Retrieved from <http://www.dpi.state.nc.us/docs/accountability/educators/fastattributes04081.pdf>
- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 22(4), 34-43.
- McMillan, J. H. (2010). The practical implications of educational aims and contexts for formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 41-58). New York, NY: Routledge Taylor & Francis Group.
- Morgan, G. A., Leech, N. L., Gloeckner, G. W., & Barrett, K. C. (2013). *IBM SPSS for introductory statistics: Use and interpretation* (5th ed.). New York, NY: Routledge.
- Nelson, M., & Eddy, R. M. (2008). Evaluative thinking and action in the classroom. In T. Berry & R. M. Eddy (Eds.), *Consequences of No Child Left Behind for educational evaluation. New Directions for Evaluation*, 117, 37-46.
- Onwuegbuzie, A. J., Dickinson, W. B., Leech, N. L., & Zoran, A. G. (2009). A qualitative framework for collecting and analyzing data in focus group research. *International Journal of Qualitative Research Methods*, 8(3), 1-21.

- Patton, M. Q. (2018). A historical perspective on the evolution of evaluative thinking. In A. T. Vo & T. Archibald (Eds.), *Evaluative Thinking. New Directions for Evaluation*, 158, 11-28.
- Phelan, J., Choi, K., Vendlinski, T., Baker, E., & Herman, J. (2011). Differential improvement in student understanding of mathematical principles following formative assessment intervention. *The Journal of Educational Research*, 104(5), 330-339.
- Popham, W. J. (2008). *Transformative assessment*. Portland, OR: Association of Supervision and Curriculum Development.
- Rubin, H. J., & Rubin, I. S. (2005). *Qualitative interviewing: The art of hearing data* (2nd ed.). Thousand Oaks, CA: SAGE Publications.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39-83). Chicago, IL: Rand McNally & Company.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: SAGE Publications.
- Shermis, M. D., & Di Vesta, F. J. (2011). Formative assessment: Using assessment for improving instruction. In *Classroom assessment in action* (pp. 83-118). Lanham, MD: Rowman & Littlefield Publishers, Inc.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1-23.
- Sprinthall, R. C. (2007). *Basic statistical analysis* (8th ed.). Boston, MA: Allyn and Bacon.
- Stake, R. E. (2010). *Qualitative research: Studying how things work*. New York, NY: The Guilford Press.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: SAGE Publications.
- Tolley, L. M. (2011, November). *Everything matters: Understanding the impact of context on formative assessment*. Paper presented at the American Evaluation Association Annual Conference, Anaheim, CA.
- Tolley, L. M. (2012, October). *But does it work? A review of teacher professional development on formative assessment*. Poster presented at the American Evaluation Association Annual Conference, Minneapolis, MN.
- Tolley, L. M. (2016). *Assessing formative assessment: An examination of secondary English/language arts teachers' practices* (Doctoral dissertation). Retrieved from SURFACE. (457)
- William, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment*, 11(3&4), 283-289.
- Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M., Tomita, M. K., & Young, D. B. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied Measurement in Education*, 21(4), 335-359.
- Zirkel, S., Garcia, J. A., & Murphy, M. C. (2015). Experience-sampling research methods and their potential for education research. *Educational Researcher*, 44(1), 7-16.

Appendix A

English/Language Arts Teachers' Experiences (ELATE) Checklist Items

Item #**Item Text**

In teaching [the concurrent enrollment course] this week (last Friday through Thursday), I...*Questioning and Discussion*

- 1 Asked students questions to determine how well they understood a concept, idea, or strategy.
- 2 Asked questions of individual students by name.
- 3 Asked questions of the class as a whole.
- 4 Asked questions specifically of students I felt were not paying attention.
- 5 Asked questions of students I thought would be more likely to respond well.
- 6 Asked questions of reticent students to help improve their participation.
- 7 Asked questions requiring brief responses (e.g., a word or phrase).
- 8 Asked questions requiring more elaborated responses (e.g., a few sentences).
- 9 Asked questions intended to stimulate a general discussion.
- 10 Used paired or small group (2-4 students) discussion to determine how well students understood a concept, idea, or strategy.
- 11 Used large group (5 or more students) discussion to determine how well students understood a concept, idea, or strategy.
- 12 Used whole-class discussion to determine how well students understood a concept, idea, or strategy.

Feedback

- 13 Reviewed and gave feedback (oral or written) on students' think/response papers.
 - 14 Reviewed and gave feedback (oral or written) on students' dialogic journals.
 - 15 Reviewed and gave feedback (oral or written) on students' portfolios of their writing.
 - 16 Reviewed and gave feedback (oral or written) on students' *draft* versions of a writing assignment.
 - 17 Reviewed and gave feedback (oral or written) on students' *final* versions of a writing assignment.
 - 18 Used a rubric or rubrics for *informal* feedback to students on their writing.
 - 19 Used a rubric or rubrics for *formal* feedback to students on their writing.
 - 20 Gave students *targeted* written feedback about their writing, such as comments on certain parts of their assignment.
 - 21 Gave students *general or holistic* written feedback about their writing, such as comments at the end of an assignment about how they did overall.
 - 22 Wrote questions when giving feedback to students on their writing to help prompt their thinking and develop their writing skills.
 - 23 Corrected students' spelling, grammar, and other mechanical errors when giving them feedback on their writing.
 - 24 Gave oral feedback to multiple students at once about their writing (e.g., discussing strategies with groups or the entire class).
 - 25 Conferenced with individual students about their writing to give them feedback and to review and discuss their work.
-