



# Navigating the complex cognitive task of classroom assessment

Helenrose Fives<sup>\*</sup>, Nicole Barnes

Montclair State University, 1 Normal Avenue, Montclair, NJ, 07043, USA



## HIGHLIGHTS

- Teachers' classroom assessment practice can be conceived of as a complex cognitive task.
- The expert teacher investigated here developed management routines and interpretation strategies to support decision making.
- Assessment management routines addressed assigning, collecting, returning, annotating, and organizing student work.
- Two families of interpretation strategies emerged: appraisal (quality) and monitoring (progress) strategies.
- Assessment management routines, interpretation strategies, and instructional decisions were symbiotic in this teacher's work.

## ARTICLE INFO

### Article history:

Received 25 September 2018

Received in revised form

17 December 2019

Accepted 12 March 2020

Available online xxx

### Keywords:

Classroom assessment

Teacher decision making

Qualitative study

Teacher expertise

## ABSTRACT

We investigated how a fifth-grade literacy teacher with classroom assessment expertise integrated assessment into his teaching practice. We employed a qualitative case study methodology to examine and document the processes this teacher used to make instructional decisions informed by classroom assessment events. We identified a repertoire of assessment relevant management routines and interpretation strategies that supported his assessment practice and instructional decision-making. Findings can be used to inform policy and organize content for teacher education by identifying a repertoire of potential management routines and interpretation strategies for engaging in classroom assessment and documenting *how* to enact these practices.

© 2020 Elsevier Ltd. All rights reserved.

Teachers are expected to know the theoretical and philosophical foundations of how to measure student learning coupled with the appropriate use of assessment related practices (DeLuca & Klinger, 2010; Volante & Fazio, 2007). In the U.S., this notion is prominent in state and national teaching standards (e.g., *Interstate Teacher Assessment and Support Consortium*, 2011) and national reform efforts (e.g., *U.S. DOE*, 2015). Despite the demand to develop teachers' assessment expertise, international findings suggest that teachers are not using assessment to inform instruction (e.g., Kippers, Wolterinck, Schildkamp, Poortman, & Visscher, 2018). This may be caused by limitations in their skills for using the information gleaned from assessment events to meaningfully inform instruction (Beswick & Sloat, 2006; Mertler, 2004). In response, policy makers and scholars have called for increased assessment learning during preservice teacher education (e.g., DeLuca & Bellare, 2013) and supported the use of structured team meetings

focusing on analyzing assessment results for practicing teachers (Mandinach & Gummer, 2013). However, instruction in classroom assessment for teachers typically occurs in a college classroom setting or in professional development events outside the classroom (Farrell & Marsh, 2016). The content examined in these experiences is typically derived from recommendations from experts in the fields of evaluation and measurement or educational psychology (e.g., Fives, Barnes, Dacey, & Gillis, 2016) rather than from successful models of teachers' *in situ* navigating classroom assessment as a complex cognitive task.

To understand the complexity of classroom assessment practices, we investigated how a fifth grade English/language arts (ELA) teacher with classroom assessment expertise engaged in assessment activities. This investigation led to a repertoire of assessment relevant practices that can be used to inform policy and organize content for teacher learning in classroom assessment.

## 1. Theoretical framework

Our theoretical framework is informed by research and theory

<sup>\*</sup> Corresponding author.

E-mail address: [fivesh@montclair.edu](mailto:fivesh@montclair.edu) (H. Fives).

from the fields of classroom assessment, cognition (specifically complex cognitive tasks), and teacher expertise. Below we describe how these fields informed and influenced our investigation.

### 1.1. Classroom assessment

Assessment is the process of obtaining “information that is used to make educational decisions about students, to give feedback to students about their progress, strengths, and weaknesses, to judge instructional effectiveness and curricular adequacy, and to inform policy” (American Federation of Teachers AFT, National Council on measurement in education NCME, & National Education Association NEA, 1990, p. 1). According to this definition assessment is not a “thing,” but instead a process that can be directed to inform teaching and learning (e.g., formative: Black & Wiliam, 2009), evaluate performance (e.g., summative/grading: Brookhart et al., 2016), or identify students’ needs (e.g., diagnostic, Brookhart & Nitko, 2006). We focus on *classroom assessment*, which situates the assessment process in individual classrooms, orchestrated by teachers (Pellegrino, 2012).

We investigate teachers’ classroom assessment practices for four reasons. First, U.S. teachers devote as much as 50% of their teaching tasks to assessment-related activities and rely largely these classroom data, as opposed to information from other standardized-type tests, to evaluate student performance and make decisions about next steps in instruction (Baird, 2010; Stiggins, 1999). Second, information derived from classroom assignments is more likely to provide teachers with meaningful representations of their students’ knowledge and skills, compared to information from standardized tests (Shepard, 1989). Third, teachers’ classroom assessment practices are a central tenet of a standards-based, accountability system of education (Earl, 2012). Fourth, teachers’ assessment practices influence the learning context, student achievement (Wiliam, Lee, Harrison, & Black, 2004), student learning (Brookhart, 2017; Wilson, 2004), and students’ identity (Bliuc, Ellis, Goodyear, & Hendres, 2011). Given this, it is not surprising that McMillan (2013) asserted that classroom assessment is “the most powerful type of measurement in education that influences student learning” (p. xxiii).

However, to effectively engage in assessment-related activities, teachers need a well-formed knowledge base of the fundamental concepts and practices in classroom assessment that are likely to positively affect student outcomes. This essential knowledge is referred to as *assessment literacy* (Gareis & Grant, 2015; Popham, 2011) and *how* teachers’ prioritize their knowledge of assessment is referred to as their *approach to assessment* (DeLuca, LaPointe-McEwan, & Luhanga, 2016). Teachers’ develop their approach to assessment over time, first through their own schooling experiences (Lortie, 1975; Pajares, 1992; Volante & Fazio, 2007); then through their course work and practicum experiences in their teacher education programs (DeLuca & Klinger, 2010; Graham, 2005; Volante & Fazio, 2007). Such experiences shape and reshape teachers’ approaches to assessment and their assessment literacy over time. Research comparing novice and experienced teachers provides evidence to support this developing assessment literacy over time. Researchers found that experienced teachers: (a) used varied assessment practices (Coombs, DeLuca, LaPointe-McEwan, & Chalas, 2018); (b) demonstrated assessment-related knowledge of theory and concepts (Mertler, 2004); (c) attached value to peer or collaborative assessment activities (Wen, Tsai, & Chang, 2006); and (d) employed a differentiated approach with regard to issues of fairness in assessment (Coombs et al., 2018).

To help guide pre-service education and inservice professional development in classroom assessment, the Joint Committee for the Standards of Educational Evaluation (2015) published the

*Classroom Assessment Standards: Sound Assessment Practices for PK–12 Teachers*. These 16 standards are organized into three areas: foundations, use, and quality. The foundations domain has six standards, which encompass the planning a teacher must do to ensure assessment practices are sound and fair. This includes ensuring that assignments have a clear purpose, align with curricular objectives, are designed to ensure that students can demonstrate their learning and are engaged in the process. In addition, the foundations domain includes standards for preparing teachers and students for assessment and communicating with stakeholders (e.g., parents/guardians) about classroom assessment activities. The next five use standards guide teachers’ use of assignment results, specifically their analysis of student work, feedback provided, and instructional follow-up. Use standards also address how assignment results are used to inform summative grades and reports of student learning. The five quality standards help teachers ensure that their assessment practices are of high quality. This refers to identifying evidence of reliability and validity as well as ensuring that assignments are inclusive of all students, fair, and free from bias. Taken together, these standards represent the knowledge, skills, and techniques teachers should and can use to collect, interpret, and use information from assessment tasks to inform their own teaching practices, and to improve student learning.

It is worth noting that the underlying purpose of classroom assessment activities is for teachers to gather evidence that will allow them to evaluate students’ knowledge, skills, and understandings. Thus, the processes of assessment that teachers engage in should allow for *valid* interpretations. Validity is a concept used in the field of educational and psychological measurement and assessment to describe the quality of judgements or decisions that can be inferred from a test score. In fact, the 1999 *Standards for Educational and Psychological Testing* stated “[v]alidity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (p. 9). However, much of teachers’ work in classroom assessment does not rely on externally created and validated tests or other single assignments. We can turn to Newton’s (2012) explanation of the “spirit of the *Standards*,” whereby he described validity as “the difference between having, or not having, a solid technical justification for using a given assessment procedure to make a certain kind of decision (for a particular group of individuals within a particular context)” (p. 25). Still this does not capture the complexity of classroom assessment, which is an ongoing process that actively integrates instruction with assessment (e.g., Brookhart, 2003; Moss, 2003). Moreover, teachers’ decisions that carry high stakes (e.g., end of year grades that lead to placement decisions) are not made based on a single test, which is the basis for the definition of validity used in the *Standards*.

Validity is important for classroom assessments, as it theoretically calls into question whether teachers’ evaluations about students’ performance are based on strong evidence. Does an “A” in a teacher’s class indicate that the student has mastered the subject or that the student has turned in all the assignments on time? There is debate as to what criteria should be used to judge validity in the case of classroom assessments as opposed to large scale assessments. For example, the expectations for ensuring validity of large scale measurement tools such as large sample sizes, multiple items, and statistical analyses, are often untenable for teachers’ classroom level assessment practices (Brookhart, 2003; Cizek, 2009; Moss, 2003; Smith, 2003). Thereby making traditional psychometric approaches of validity to classroom assessment practices equally untenable. In the present investigation we explored the ways that one teacher balanced the need to make valid evaluations with his overall goals for teaching and supporting his students’ learning. We

identify instances where he demonstrated concern for validity in terms of response content (is the material assessed what was supposed to be assessed) and response process (is the material assessed at the targeted cognitive levels; AERA, APA, NCME, 1999).

### 1.2. Classroom assessment as a complex cognitive task

A complex cognitive task refers to the “integration of skills, knowledge, and attitudes, and the extensive coordination of constituent skills in new problem situations” (van Merriënboer, Kirschner, & Kester, 2003, p. 6). Engaging in classroom assessment can be understood as a complex cognitive task as it requires teachers to integrate their knowledge of content (Feiman-Nemser & Parker, 1990), pedagogical content knowledge (Shulman, 1986), assessment (Fives & Barnes, 2017), the needs of their students (Wiggins, 1998), and the contexts of the school, classroom, and community (Villegas & Lucas, 2002). Moreover, teachers often make determinations about students’ grades and achievement based on multiple sources of information including varied assessment events (e.g., tests, homework, essays; Black, 2012), informal observations of students (Allal, 2013), and perceptions of student effort (Brookhart, 1994). Allal’s (2013) qualitative analysis of teachers’ processes for determining end-of-semester grades illustrated the complex nature of this assessment task, as these teachers pooled multiple types of student work with their knowledge of students’ lives (e.g., family illness) to make decisions about assigned grades.

Successful completion of complex cognitive tasks requires individuals to draw on their understanding of the discipline, while executing a variety of cognitive strategies and routines (van Merriënboer et al., 2003). Cognitive strategies are a form of procedural knowledge (i.e., knowing *how* to do something; Alexander, Schallert, & Hare, 1991) that individuals employ to help them on goal directed tasks (MacArthur, 2012). This may include identifying the plot in a story or outlining an argument in an essay. In addition to cognitive strategies, Leinhardt, Weidman, and Hammond (1987) argued that complex cognitive tasks require the implementation of routines. Routines are “small cooperative scripts of behavior used to support several activity structures, for example, passing out papers” (Leinhardt et al., 1987, p. 135). Individuals form routines by recognizing that certain patterns of behavior or procedures are an appropriate response to similar tasks. Regularly executing a routine can lead to automatization, freeing up cognitive processing space for individuals to attend to other demands (Lampert, 2009).

Despite the benefits of having and using a repertoire of routines and cognitive strategies, researchers have argued that pre-service teachers do not have the cognitive strategies to differentiate among various assessment purposes and types (Maclellan, 2004), and are varied in their ability to generate and use assessment strategies effectively without instruction (Volante & Fazio, 2007). More recently, Fives & Barnes (2017) found that pre-service teachers generated their own strategies for selecting test items, in the absence of formal instruction. These self-generated strategies were less effective in helping them to choose items with sound validity evidence.

In contrast, studies of practicing teachers have revealed that teachers developed routines and strategies that they modified and adapted in practice as needed to support their classroom assessment practices (e.g., Allal, 2013). In general, however, these routines and strategies have been underspecified in empirical research presentations, which has subsequently masked the complexity of teachers’ work in this area. The implication then is that teacher educators do not have a set of empirically-grounded routines and strategies with clear examples that can be explicitly taught to novices. Therefore, descriptive research on such specific routines

and strategies is needed.

### 1.3. Teachers’ expertise

One approach to understanding the complex processes teachers use to manage their classroom assessment tasks is to look closely at the practices and conditions of those recognized as experts in the phenomenon (Palmer, Stough, Burdinski, & Gonzales, 2005). Teachers with expertise may have salient craft knowledge,<sup>1</sup> in particular the declarative (i.e., knowing *what* or *that*; Alexander, Schallert, & Hare, 1991) and procedural knowledge of strategies and routines that can be accessed and understood as exemplars for practice (Grimmett & MacKinnon, 1992). Across domains of practice, experts are those who have knowledge that is vast, deeply integrated, and that can be easily accessed; they are able to select and use appropriate strategies in context (Ericsson & Smith, 1991; Glaser & Chi, 1988). Experts and expert performance have been examined in many fields, including teaching, to identify learning trajectories and facilitate expertise development (e.g., Bereiter & Scardamalia, 1993; Bransford & Schwartz, 1999), to provide models of successful learners (e.g., Pellegrino, Chudowsky, & Glaser, 2001), and to identify problem-solving and decision making strategies to teach to learners in a domain (e.g., Harris & Graham, 1992; VanSledright, 2002).

In the context of teaching, researchers found that expert teachers used a host of cognitive strategies (e.g., organizing a lesson into action segments) and routines (e.g., homework checks) that they executed frequently and flexibly in their classrooms (Leinhardt & Greeno, 1986). Expert teachers demonstrated qualitatively different strategies than novices when presented with the same classroom management teaching scenarios (Kim & Klassen, 2018). Furthermore, expert teachers focused on the underlying principle evident in scenarios whereas novices attended to surface-level considerations (Kim & Klassen, 2018). With regard to routines, expert teachers ensured that students were aware of and able to carry out classroom routines with little explanation or monitoring (Leinhardt & Greeno, 1986). In contrast, novice teachers were less skilled at implementing routines, frequently changing their patterns of behavior and neglecting to *teach* students how to execute routines properly. Routines are not just about executing steps or procedures; equally important is understanding the context and professional norms to know when, where, and how to use routines appropriately (Feldman & Pentland, 2003). Therefore, expert teachers may be a source for identifying highly complex routines developed from a deep and relevant understanding of the context and work of teaching.

## 2. Rationale for a case study of expertise in classroom assessment

Drawing from the fields of organizational theory and the sociology of work, Warren Little (2012) identified the importance of close analysis of teachers engaged in classroom assessment practices to expose “aspects of practice that otherwise remain opaque” (Warren Little, 2012, p. 144). Alm and Colnerud (2015) noted that a close examination of how teachers enact assessment practices is both desirable and difficult. In this investigation we took up these challenges with the goal to expose the work of teachers, make explicit complex cognitive tasks involved in their work around

<sup>1</sup> Craft knowledge is understood as teachers’ ability to bring “all of the knowledge bases to bear on the act of teaching” indicating an occupational “savvy” to engage varied bodies of knowledge in specific contexts for defined goals that is developed through experience and reflection (Grimmett & MacKinnon, 1992, pp. 387–388).

classroom assessment, and to explore the strategies and routines developed to address these tasks.

To explore how teachers integrate assessment into their teaching practice, or the strategies and routines that they use or develop for their professional contexts we chose to conduct a qualitative case study of one teacher's engagement in classroom assessment activities. Case study research allows for a deep exploration into a bounded system in order to describe phenomena of interest (Bogdan & Biklen, 1997; Merriam, 1998; Stake, 1995; Yin, 2017). Case study descriptions provide examples beyond the personal experiences of individuals and may expose, for the reader, what is known and unknown in their own schema. A major strength of case study research is that it allows researchers to study the phenomenon of interest within the real-world context in which it occurs (Yin, 2017). Importantly, the descriptive nature of case study research allows readers into the world of the case that may be distinctly different or similar to one's own experiences (Bogdan & Biklen, 1998). Yin, 2017 argued that case study research is warranted when three conditions are met: (a) when researchers have how or why research questions, (b) there is limited (if any) control of behavioral events, and (c) the phenomenon of study is a contemporary issue. The present investigation meets all three of these conditions. We are interested in *how* teachers use information from assessment to make sound instructional decisions, as researchers we have no role in directing teachers' work, and the issue of classroom assessment practices is as we described in the introduction of current importance in the US and abroad.

As a qualitative investigation using the case study design we do not argue that the findings from this teacher are representative of a larger teaching population or that they would "generalize" to other teachers. Rather, we follow the recommendations from Guba (1981) and Lincoln and Guba (1985) to establish trustworthiness of our inquiry as a means ensuring rigor in our work. *Trustworthiness* is used to establish that the findings emerging from the research process can be used to inform theory, research, and practice. Within this paradigm a monolithic truth that generalizes to all populations does not exist. There are, instead, many truths that exist within individuals situated in physical and temporal contexts. Thus, under the banner of trustworthiness, qualitative researchers concern themselves with issues of credibility, transferability, dependability, and confirmability, which we adhere to in our methodology. Cognitive functioning of adults engaged in the work of classroom assessment can inform theory, research, and practice if those complex cognitive tasks can be identified, described, and situated in a context that allows the reader to transfer the learning to their own experience.

We chose to investigate the classroom assessment practices of a teacher who demonstrated hallmarks of expertise as described by the Model of Domain Learning (MDL; Alexander, 1997; Barnes & Fives, 2018). Howley, Howley, Henning, Gilla, & Weade, 2013 argued that investigators of teachers' knowledge and practice should take an *assets perspective* that seeks to learn from teachers the nature of their practice. However, "not all teachers possess equal amounts of knowledge or the same knowledge about assessment." (Howley et al, 2013, p. 28). Therefore, it is reasonable for researchers to seek out teachers who demonstrate expertise in the phenomenon of interest.

According to the MDL hallmarks of expertise include high levels of knowledge, interest, and strategic processing around classroom assessment (e.g., Alexander, 1997). We recognize expertise as including the ability of teachers to select, use, and generate appropriate strategies and routines for the context of their work. Thus, this targeted investigation allowed for an in-depth understanding of many potential routines and strategies aligned with specific goals or targets identified and explained by the teacher. We

forwarded the following research question: How does a fifth grade English/language arts (ELA) teacher with classroom assessment expertise manage the complex cognitive tasks associated with classroom assessment practice?

### 3. Method

We engaged in an instrumental intrinsic qualitative case study analysis to garner an in-depth understanding and detailed description of the phenomenon of managing the complex cognitive task of classroom assessment as demonstrated in the practice of one teacher (Miles & Huberman, 1994; Stake, 1995). According to Stake (1995) a case is *instrumental* when participants are selected to gain insight into a specific phenomenon, in this case classroom assessment practices. Thus, the goal of this instrumental case was to understand how teachers engaged in classroom assessment tasks as part of their daily teaching. An *intrinsic* case refers to the investigation of a particular case (e.g., teacher, school, group) that demonstrates unique or special qualities (Stake, 1995). In this paper, we explored the practice of Mr. Sparrow,<sup>2</sup> because as we analyzed our data for a larger study of seven teachers (Barnes, Fives, Mabrouk Hattab, & SaizdeLaMora, 2020) it became apparent that Mr. Sparrow engaged in unique classroom assessment practices that supported his teaching practice. Specifically, he demonstrated expertise in classroom assessment beyond other teachers in our sample and was also able to articulate his thinking in elaborate ways. We sought to explore and understand his practice to expose the complex work of teaching with respect to classroom assessment.

#### 3.1. Participant and participant selection

Mr. Sparrow, was a thirty-four year-old male, fifth grade (age 10) ELA teacher with 12 years of teaching experience. Mr. Sparrow was White and held a master's degree plus 30 additional graduate level credits in education/teaching-related courses. At the time these data were collected the state reported that of the 114,869 certified teachers 23% were male and 84% of all teachers identified as White. In his district of 374 teachers, 77 were male (21%) and 93% identified as White. Of the 30 teachers in his school three were male and all of the teachers identified as White. As such we recognize that Mr. Sparrow is not reflective of the overall teaching population in this state which is primarily female. However, we selected him for this investigation as an *intrinsic* case of expertise in classroom assessment, this is described below (Stake, 1995).

We selected Mr. Sparrow as a participant for this study based on four criteria: First, we used Palmer and colleagues' (2005) two-gate process for identifying expertise. Gate 1 required that the teacher have a minimum of three years of experience in his present context and requisite degree(s) and certification(s) for his present position. Gate 2 required recognition of expertise from informed perspectives (e.g., principal) relative to salient indicators. Using a nomination form describing the goals of the study and the Gate 1 criteria we asked school leaders to nominate teachers with expertise in classroom assessment (Gate 2). Mr. Sparrow's principal nominated him to be considered as a participant in this study.

Second, we interviewed Mr. Sparrow by phone about his assessment practices. The protocol included questions such as, "What is your best assessment 'trick' or strategy that you think every new teacher should learn? Why?" and "How do you use data for instructional improvement?" Mr. Sparrow's responses were evaluated using a rubric comprised of the three components of the

<sup>2</sup> All names are pseudonyms.



MDL (i.e., knowledge, strategies, and motivation). Mr. Sparrow was rated the highest level, “expertise” on all evaluative criteria.

Third, we considered performance indicators that are typically used by school administrators to determine teaching competence. In Literacy and Mathematics, 95% and 93% of Mr. Sparrow's students scored at Levels 4 or 5 (advanced proficient) on the Partnership for Assessment of Readiness for College and Careers (PARCC) exam, a required assessment for all public schools in the state where this study was conducted. These scores compared to a US national average of 40.5% and 36.2% respectively. Moreover, Mr. Sparrow was one of 92 teachers out of the 392 teachers in his district to be rated “highly effective<sup>3</sup>” in each domain of the educator evaluation system rubric. Fourth, after comparing all the teachers who were part of our larger sample of seven on their assessment practices, we selected Mr. Sparrow as a distinct example of a teacher with expertise in classroom assessment (Barnes, Fives, Mabrouk Hattab, & SaizdeLaMora, 2020).

### 3.2. Context

Mr. Sparrow taught in a middle-class town (average family household income: US\$148,500) that included White (68.2%), Asian (23%), and Hispanic and Black (7%, combined) residents. On average, the school enrolled 343 students per school year, and had a faculty to student ratio of 1:9. According to the publicly available school report card, none of the students were identified as economically disadvantaged, 2.6% of students had limited English proficiency, and 11.1% were classified and received special education services. Mr. Sparrow's fifth grade class included 24 students (13 girls and 11 boys).

When entering Mr. Sparrow's classroom, there was a round table where he worked and conferred with students. There was also a reading center with a large blue carpet, flipboard, and rocking chair where he held mini-lessons. Desks were arranged in two long rectangular tables with students on either side of the table. There was a SMART board, whiteboard, and brown writing board at the front of the room. Laptops, shared with the other fifth grade class, were available for the students to use regularly.

Mr. Sparrow taught ELA using The Teachers' College Reading and Writing Workshop curriculum, which emphasizes skill instruction, independent reading of self-selected texts, mini-lessons, and frequent conferencing with students (Calkins & Tolan, 2010). At the start of each marking period, Mr. Sparrow formed groups of 2–3 students, and each group selected a “just right book” based on their common interests and reading level. Mr. Sparrow used multiple forms of assessment tasks to evaluate students' literacy skills. For clarity in writing our findings, we used four of his common assessment sources: Post-it assignments, Reading Responses, Personal Narrative (in Google docs), and Conference Notes from which to draw examples. Here we provide a brief description of these assessment tasks.

Mr. Sparrow used Post-its (small square sticky notes) during instruction to both check students' understanding of the text or their progress in applying a particular reading strategy/skill as well as to help students stay engaged with their reading. In the excerpt below Mr. Sparrow described how he explained this assignment to his students:

About the first two or three weeks of our reading workshop is also spent on how we communicate what we are learning and what are seeing. So ... we talk about why we write on Post-its in the first place - so that we can stay engaged in our books but at the same time, we can show the skills and strategies that we have and we can ask the questions that we need to (II, 10/29).

For each Post-it assignment, Mr. Sparrow posed a short task/question to the class (e.g., Write a prediction about your character based on your character theory) and students recorded their responses on a Post-it (see Fig. 1). Each Post-it task had the same required elements: skill/strategy used, ideas supported with “because,” text evidence, accurate ideas/connections to the book, and page number. These elements are detailed in the Perfect Post-it rubric (see Fig. 1), which was designed by Teachers' College.

In Fig. 1 we see that the student identified the skill, prediction, and that s/he does indeed make a prediction about whether or not the character will stop at the town, using information that is accurate for this book. However, there is no use of “because” to explain the reason for the prediction and no text evidence to support the prediction, therefore this would be evaluated as a 3 Star Post-it (in the findings section under appraisal strategies we share Mr. Sparrow's analysis of this Post-it). While at the beginning of the school year Mr. Sparrow routinely assessed students' use of the required elements of the Post-its with the rubric and recorded scores, once students were comfortable with the structure of the task (by the second marking period) Post-its were used informally to provide Mr. Sparrow with information on students' progress and were not given a summative score.

Students wrote a weekly Reading Response (see Fig. 2) in which they summarized their reading from their just right book and practiced using the week's reading strategy/skill. Each Reading Response followed the same structure: students provided background information about the text (e.g., title), evidence for and a description of how they used that week's reading strategy (e.g., activate prior knowledge), and a conclusion and prediction for what they thought would occur next. Several different strategies were used to evaluate students' Reading Responses and students received a recorded/summative score for each Reading Response. Students completed Reading Responses on a weekly basis and these assignments played a central role in Mr. Sparrow's literacy instruction and assessment. For the Reading Response assignments, the target skills were based on the reading skill or strategy students learned that week.

Mr. Sparrow scored students' Reading Responses using a rubric developed by Teacher's College (see Fig. 3). The rubric was an analytic rubric with seven evaluative criteria (heading, evidence of reading, evidence of reading strategy, thorough explanation of reading strategy, conclusion, prediction, and spelling, punctuation, grammar) scored on a 4-point scale. Students received an overall letter score which was determined by averaging their score on each evaluative criterion and referencing the key at the bottom of the rubric.

In the writing component of the ELA curriculum students were required to write a personal narrative. Mr. Sparrow used the shared folder/document abilities in Google docs to collect and provided feedback on student work in progress. The assignment had a rubric used to evaluate it, however we only observed him assessing initial drafts of this work.

Mr. Sparrow routinely met with either individual students or small groups of students in reading/writing conferences. During these teacher-led conferences he made observations of students' skill level and drafted conference notes about each student. The conference notes were kept electronically in a word processing

<sup>3</sup> The Teacher Evaluation System in this district is based on a modified version of the Marshall Teacher Evaluation Rubric (<https://marshallmemo.com/articles/Marshall%20Teacher%20Eval%20Rubrics%20Aug.%2031,%2011.pdf>). It includes six domains (Planning and Preparation for Learning, Classroom Management, Delivery of Instruction, Monitoring, Assessment, and Follow Up, Family and Community Outreach, and Professional Responsibilities) and teachers are rated on a scale from 1 (does not meet standards) to 4 (highly effective).

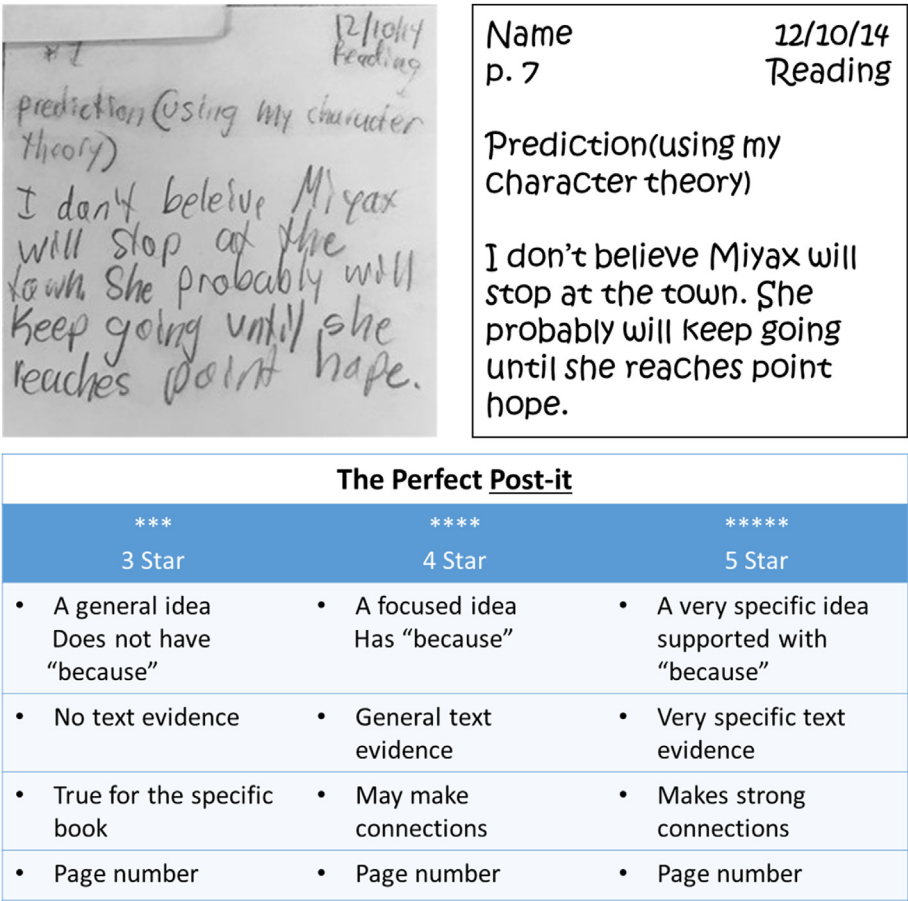


Fig. 1. Sample post-it assignment and rubric.

program. The format for these notes included four columns: date, reader/partnership/club, compliment, and coaching point. Each week he created a new document for the conference notes so that he could go back through the notes systematically.

3.3. Data collection timeline

We gathered data for this investigation primarily through interviews with Mr. Sparrow in his classroom. We conducted an initial interview (II) with Mr. Sparrow in October. Then in December across a two-week period we conducted four classroom observations of literacy instruction. During this time we also conducted six intensive think-aloud interviews (TA; one in January due to weather and school holidays). At the end of the school year, in June, we conducted a closing interview (CI). Our overall target time period for collecting the classroom observations and TA interviews was from November 1– March 1. We wanted to be far enough into the school year so that teachers had time to establish their classroom routines and develop an understanding of their students. We also needed to have data collection completed well before April when the state-wide standardized testing occurred. .

3.4. Data sources

3.4.1. Classroom assessment think-alouds (TA)

Typically, TA protocols are used to access the cognitive processes of experts engaged in problem solving tasks (Ericsson, 2006). Thinking aloud involves reporting whatever thoughts are available

to the teacher in-the-moment (Pressley & Afflerbach, 2012), which can afford opportunities to better identify potential cognitive processes that underlie his/her attempts to understand student work and use it to make decisions (or not). During the two-week cycle, two members of the research team coordinated with Mr. Sparrow to be present whenever he engaged in instructional planning and/or any classroom assessment activity. We introduced Mr. Sparrow to the concept and goals of the think aloud interview and then, following our protocol (see Fig. 4), prompted him to think aloud while he performed those tasks. Specifically, we asked him to read the work aloud (when appropriate) and describe his thoughts and reactions to the work, we encouraged him to elaborate and make his thinking explicit. He frequently commented on seemingly tangential issues while engaged in his evaluation of student work. We encouraged this kind of free-flowing thinking through all of the TA interviews. These sessions were audio-recorded and transcribed. The interviews totaled 241 min, averaging 40 min each, the shortest TA was 22 min and the longest 52 min.

3.4.2. Classroom observations

We observed Mr. Sparrow's ELA class for four consecutive days. The purpose of the observations was to get a sense of his classroom routines and structure as related to classroom assessment practices. We gathered detailed field notes focused on assessment-related activities and classroom routines. Each observation was conducted by two researchers using a face-to-face observation protocol that included a session overview and running record. In total we observed Mr. Sparrow for 12 h and 25 min.

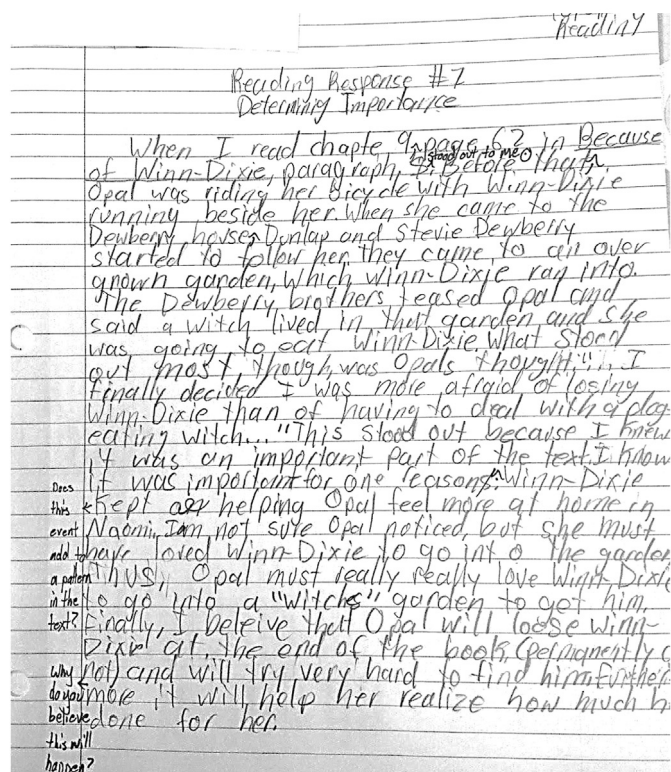


Fig. 2. Sample reading response.

### 3.4.3. Classroom assessment interviews

We conducted an initial (II; 96 min) and a closing interview (CI; 70 min) with Mr. Sparrow focused on his assessment practices. These questions focused on his overall perspective on assessment (e.g., Explain your assessment system to me as if I was a parent.), the nature of the assignments he collected or had access to (e.g., What assignments do you use that result in information about student learning?), his assessment analysis processes (e.g., What do you do with an assignment once it is collected?), and his perceptions of his professional context around assessment (e.g., In what ways do you feel you have (or do not have) the authority to make decisions based on information from your classroom assignments?). In addition, we used the CI to engage in member checking of our in-progress analysis of data gathered during the Observations and TAs from earlier in the year.

### 3.4.4. Material artifacts

At all data collection points we gathered material artifacts to support our understanding of Mr. Sparrow's assessment use. This included digital copies of classroom assignments, records of student data, planning materials, and any other related instructional/assessment-related materials.

### 3.5. Thematic analysis

We engaged in a thematic analysis of these data and conducted recursive emergent coding to answer our research question (Miles, Huberman, and Saldana, 2014). This included a two-phase process of ongoing and recursive interrogation of the research data

gathered. Phase One included three procedures: Listening Guide Protocol (e.g., Brown & Gilligan, 1992), initial case development (LeCompte, Preissle, & Tesch, 1993), and peer debriefing. These analyses resulted in analytic memos and analysis artifacts that guided ongoing reflection, data collection, and initial analysis ideas for Phase Two. The Phase Two analysis occurred after data collection was complete. All data were entered into NVivo for coding and analysis.

Analysis in Phase Two was directed at answering our research question by focusing on the strategies and routines Mr. Sparrow employed to facilitate his assessment activities. We first determined initial codes for each independent action, thought, or idea related to classroom assessment performed by Mr. Sparrow. Code development was thorough and recursive. We conferred frequently with one another to discuss terminology and definitions and added new codes as needed to best capture the data.

Query reports of the data by code helped us to identify Mr. Sparrow's cycle of assessment practices, which we initially attempted to represent as visual strategy maps. However, as we continued to change and refine these maps the recursive, iterative, and symbiotic nature of Mr. Sparrow's strategy-use, routines, and decision making emerged as central to our findings. Rather than a linear or circular process map, we found Mr. Sparrow's practice to include strategies that were interwoven with routines. We discuss this further in the results section. Thus, to capture and better represent Mr. Sparrow's assessment activities, we grouped related strategies and routines.

Trustworthiness of the analysis was addressed by triangulating multiple sources of research data (Creswell & Miller, 2000; Lincoln & Guba, 1985). During research data collection and analysis we searched for negative instances of potential patterns or alternative explanations that could help us interpret the research data. Lastly, member checks were conducted with the teacher both during the CI and via email correspondence.

## 4. Findings: assessment relevant practices

Within the context of Mr. Sparrow's assessment activities we identified practices that supported his instructional decision making. *Instructional decisions* were directed at issues of who (individuals, groups, class), what (content/skill), when (right away, next day), and how (method of instruction/interaction). Mr. Sparrow came to these decisions through his use of assessment management routines and interpretation strategies. *Assessment management routines* included the methods he used to assign, collect, return, annotate, and organize student work. Mr. Sparrow used *interpretation strategies* to make meaning of student work. In the next sections we describe the assessment relevant practices that emerged and highlight the instructional decisions that were embedded in these practices or occurred symbiotically with them. To help the reader ascertain the specific data sources for each excerpt we labeled them using this format: data source (II: Initial Interview; TA: think-aloud, CI: Closing Interview), and the date.

### 4.1. Assessment management routines

We identified five sets of assessment management routines that describe how Mr. Sparrow systematically assigned, collected, annotated, organized, and returned student work. Table 1 offers a description of the purposes of the routines and examples.

#### 4.1.1. Routines for assigning student work

Mr. Sparrow used routines for assigning student work to support students' ability to complete assignments with ease. These routines ranged in complexity from a homework board in the front



	1	2	3	4
Heading	The reader has started the response with a partial heading, which includes: name, number, date, subject, response number, and strategy. Three or more elements of the heading may be missing.	The reader has started the response with somewhat of a properly constructed heading, which includes: name, number, date, subject, response number, and strategy. Two elements of the heading may be missing.	The reader has started the response with most of a properly constructed heading, which includes: name, number, date, subject, response number, and strategy. One element of the heading may be missing.	The reader has started the response with a properly positioned and constructed heading, which includes: name, number, date, subject, response number, and strategy.
Evidence of Reading	The reader has shown some evidence of reading in partially providing the following: book title, chapter number, page number, characters, quote, and summary.	The reader has shown evidence of reading in providing much of the following: book title, chapter number, page number, characters, quote, and summary.	The reader has shown substantial evidence of reading in providing most of the following: book title, chapter number, page number, characters, quote, and summary.	The reader has shown thorough evidence of reading in providing all of the following information: book title, chapter number, page number, characters, quote(s), and summary.
Evidence of Reading Strategy	The reader has partially identified the reading skill or strategy used as s/he read. Clarification is necessary for accurate identification of that reading skill/ strategy.	The reader has identified the reading skill or strategy used as s/he read. Clarification may be required for better identification of that reading skill/strategy.	The reader has identified the reading skill or strategy used as s/he read. Very little clarification may be required for better identification of that reading skill/strategy.	The reader has clearly identified the reading skill or strategy applied as s/he read.
Thorough Explanation of Reading Strategy	The reader has demonstrated partial proficiency in applying the reading skill/ strategy. The reader needs to clarify his or her understanding of the skill/strategy, as well as the way in which that skill/ strategy was applied.	The reader has demonstrated some proficiency in applying the reading skill/ strategy. Some clarification may be needed in relation to either the reader's understanding of the skill/strategy, or the way in which that skill/strategy has been applied.	The reader has demonstrated proficiency in applying the reading skill/strategy. Slight clarification may be needed in relation to either the reader's understanding of the skill/ strategy, or the way in which that skill/ strategy has been applied.	The reader has clearly demonstrated true proficiency in applying the reading skill/strategy. No questions exist as to either the reader's understanding of the skill/strategy, or the way in which the reader applied that skill/strategy.
Conclusion	The reader has composed a conclusion which ties some elements of the response together.	The reader has composed a conclusion which ties many elements of the response together.	The reader has composed a conclusion which ties most elements of the response together.	The reader has composed a clear conclusion which thoroughly ties all elements of the response together.
Prediction	The reader concludes by creating a prediction which requires further explanation.	The reader concludes by creating a prediction consisting of both the initial prediction and explanation thereof. One element of the prediction may require further explanation.	The reader concludes by creating a prediction consisting of both the initial prediction and an explanation thereof.	The reader concludes by creating an insightful prediction consisting of both the initial prediction and a clear explanation thereof.
Spelling, Punctuation, Grammar	The reader has composed a response with 7 or more errors in spelling, punctuation, or grammar.	The reader has composed a response with 3 - 6 errors in spelling, punctuation, or grammar.	The reader has composed a well written response with only 1 - 3 errors in spelling, punctuation, or grammar.	The reader has composed a very well written response with no errors in spelling, punctuation, or grammar.

Scale: Below 1=N, 1.0-1.9=P, 2.0-2.9=G, 3.0-3.6=VG, 3.7-4.0=E

Name: \_\_\_\_\_ Avg.: \_\_\_\_\_ Letter Grade: \_\_\_\_\_

Fig. 3. Reading response rubric.

of the classroom where all assignments were posted, to establishing a format for recurring assignments. He stocked the desks/tables in the classroom with baskets of Post-its and writing tools so students could readily complete assigned tasks, and assignments were due on standard days each week (e.g., Reading Responses were due on Tuesdays). Mr. Sparrow used the same format for the Post-it and Reading Response assignments throughout the year, changing the target skill, yet retaining the overall organization of the assignments, allowing for consistent expectations.

Routines regarding how to complete recurring tasks were taught at the beginning of the year and examples were shown throughout the school year to model successful performance. Mr. Sparrow explained that he regularly searched for student examples that could serve as models to share with the class. This led him to engage in instructional decision making. For example, while interpreting a student's Post-it for the use of character theory, predication, and its overall pattern, he stated:

So, I would be tempted to share this with the class, so whenever we see something that's a perfect three [rubric score], I put it on the smartboard and we go over it as the standard to work toward (TA-12/10).

In this particular example, Mr. Sparrow contemplated using this student's work as a model (i.e., a *what* decision), and also described *how* he would do this. In order to select this model, Mr. Sparrow had to have a clear understanding of the components of quality work for himself, how each component related to the parts of the task, and how this particular model illustrated those components. Mr. Sparrow needed assessment information that aligned with his

specific daily objectives and overarching alignment with the content standards (Cizek, 2009) so that he could make sound instructional decisions for his students. This shows initial evidence that Mr. Sparrow had a conceptual understanding of the need for validity evidence based on test content (AERA, APA, NCME, 1999; Newton, 2012).

In addition, Mr. Sparrow presented and explained the model to students to help them form a better understanding of the components of successful performance and how to use these components in their responses. When students understand of the markers of successful performance this can help them make sense of feedback (Hounsell, 1997). Once Mr. Sparrow and his students developed a clear understanding of the expectations, they became *shared* expectations. Mr. Sparrow used these shared expectations as a goal post to make comparisons between particular student work and the model, and to facilitate communication with students on how to improve their work. Furthermore, because this assignment was recurring, the components remained the same, allowing instructional decisions and conversation to focus on progress and growth. Throughout the year Mr. Sparrow continued to look for examples of student work to share with the class as models, thereby reinforcing this assignment routine, which facilitated his instructional decision making.

#### 4.1.2. Routines for collecting student work

Mr. Sparrow used routines for the *collection* of different types of student work. For instance, exit slip Post-its (graded) were collected in one colored basket and process Post-its (ungraded) in a different colored basket. Mr. Sparrow assigned each student a number using the alphabetical order of the students' last names.



## Data Use Think-Aloud Interview Protocol

### *Orientation to Think-Aloud Interview*

Hello, thanks once again for agreeing to participate. As you know the goal of this study is to identify the overall process you engage in and the specific, microprocesses you use to convert student responses into information that informs your practice.

During the next two weeks I will be working with you to develop an understanding of your thinking and reasoning around using classroom level data to inform your teaching practice.

We have scheduled these times to coincide with when you are typically interacting with students' classroom performance data and/or planning for instruction.

### *Instructions/Reminders*

During these sessions, I am going to ask you to engage in your work as you would normally, and that you think aloud while doing so. The goal is for you to make your inner voice audible to me.

- Don't try to plan what to say or speak after the thought, but rather, let your thoughts speak, as though you were really thinking aloud.
- In order to understand your thinking processes, especially in the beginning of the session, it would be helpful if you could explain each step of what you are doing.

I will try to refrain from interrupting your flow, but may remind you to talk aloud, or upon occasion ask for further explanation of a decision or process I observe.

While engaging in this interview, I will take notes of what you are saying and make copies of materials you use (with your permission and with any student information removed). But I will rely primarily on the audio recording of your think-aloud process.

As you discuss the assessment data you may want to talk about a particular student or refer to his/her work. If that occurs, please do NOT say the student's name, and use one of these post-it notes to cover any identifying information. I am interested in the meaning you make from the student data, not the students.

Please remember that I am not here to evaluate or critique you, but to understand your process. The more you can talk about that process the more helpful it will be for my research. If you need or want a break just let me know.

Do you have any questions?

<<<<respond to appropriate questions>>>>

Let's get started.

### *Closing*

Thank you once again for your time and willingness to share your work with us. While we have scheduled these interviews to coincide with your schedule we realize that you may engage in some planning or assessment related activities when we are not present. Thus, we are lending you this audio recorder and we ask that, should you do any planning or assessment activities when we are not present, that you continue to think-aloud during this time and to record your process for us.

Fig. 4. Think-aloud protocol.

Students included this number on all paper assignments and Mr. Sparrow collected student work in this order. Similarly, he programmed Google Docs to organize students' work to appear in alphabetical order. He described two reasons for this routine. First, was the basic need to simplify recording students' scores in his gradebook, which was set up alphabetically. Second, when asked if papers were graded in any particular order he replied:

Typically alphabetically, because what I do *not* want to do is get into a pattern where I am reading either the strongest – what I

think will be the strongest essays first or essays that really need the most work first. So I like an even disbursement, and this year, alphabetically works. Last year my strongest writers were in the beginning of the alphabet and it took me a while to realize this but the first six students were very strong writers, so when I got to that seventh student, I was pretty harsh on the grading because by comparison, you know, I had seen some good things. So this year we are pretty evenly dispersed so I try to keep it that way (TA-12/11).

**Table 1**  
Assessment management routines.

Routines	General Descriptions	Examples of Specific Routines
Assign	Routines selected and implemented to facilitate when and what kinds of assessment tasks were assigned to students	<ul style="list-style-type: none"> <li>Homework board posted in same place each day</li> <li>Reading Responses were due every Tuesday</li> <li>Reading Response and Post It assignments followed same format</li> <li>Bins with Post Its were at each student table</li> <li>Exit Slip Post Its had the same format as regular Post its. However these were intended for the student to show an example of his/her best work or practiced skill</li> </ul>
Collect	Routines for when, where, and how assessment tasks were gathered from students.	<ul style="list-style-type: none"> <li>Colored baskets for different Post It assignments (i.e., process vs. exit slip)</li> <li>Collected assignments in number order</li> <li>Conference notes forms and files had the same format</li> <li>Google docs procedures were consistent</li> </ul>
Annotate	Routines for marking student work that convey information to himself and/or the students.	<ul style="list-style-type: none"> <li>For students               <ul style="list-style-type: none"> <li>Doing well: ☺, ✓, comments</li> <li>Yellow highlight indicated a spelling, punctuation, or grammatical error</li> </ul> </li> <li>For teacher: Skill level: 0 (beginning);/(developing); + (secure); Shorthand</li> <li>Both               <ul style="list-style-type: none"> <li>"See Me:" prompted student to seek help and for teacher to note the part of the text that needed discussed</li> <li>* with [text in brackets]</li> <li>Google doc comments</li> </ul> </li> </ul>
Organize	Routines for putting assessment tasks and information inferred from student in some logical order for immediate and/or future use.	<ul style="list-style-type: none"> <li>Google portfolios</li> <li>File cabinet of student files</li> <li>Skill development chart/file</li> <li>Sorting</li> </ul>
Return	Routines for giving students assessment tasks back to students and families.	<ul style="list-style-type: none"> <li>Google docs (process feedback)</li> <li>Friday folders</li> <li>Student papers were organized in order of conference priority for return</li> </ul>

This routine for collecting student work also supported his ability to better interpret student work. While unstated, Mr. Sparrow's comment about the order of scoring papers is reflective of concerns regarding intra-rater consistency (Cizek, 2009). *Intra-rater* consistency refers to the degree to which a single evaluator consistently applies the same criteria across a number of instances of student work. In this case, how well does Mr. Sparrow apply the same expectations for quality across each of the reading responses he scored? In the comment above, Mr. Sparrow reported that in the previous school year he realized that the order in which he read student papers had an effect on the leniency or severity of his scoring. Specifically, alphabetical listing of students, by chance, had six strong writers at the top of his list. Thus, he identified a concern about his own intra-rater consistency. Awareness of this concern allowed him to evaluate his routine in the current school year to ensure that the order in which he scored papers did not allow him to establish a schema for quality work based on students who were particularly strong or weak, but rather ensured a mix of potential work quality. Thus, not only did Mr. Sparrow engage in collection routines to expedite the management of papers, but he also recognized the potential weaknesses of such routines for interpreting student work and that this could change with each class. Thus, routines were not mindlessly followed, but re-examined throughout his practice and sensitive to his context.

#### 4.1.3. Routines for annotating student work

Mr. Sparrow engaged in a series of routines to systematize his annotation of student work. These routines were tied closely to his interpretation strategies (described in the next major section). We classified them as routines because of the way these practices were repeatedly engaged to communicate to both his students and himself about the nature of their work.

Mr. Sparrow used annotations such as smiley faces, check marks, and brief comments to communicate to students about their progress. Mr. Sparrow explicitly taught students what each annotation indicated. Earlier in the year he dedicated substantial time to model for students how to interact with Google Docs and explain

what his various annotations and comments meant. He described systematic annotations for the majority of his feedback to students. For instance, Mr. Sparrow used the comment and highlighting features as key annotation routines for tasks completed in Google Docs:

I will do one of two things, I'll make comments, which highlights in tan. Or I will highlight something in yellow if it is a spelling, punctuation or error in grammar. Then what happens is, it will be up to them. They will have to figure out what needs to be changed (TA-12/11).

Comments on students' writing assignments typically supported or directed students to elaborate or clarify their writing (e.g., "consider revising," TA-12/11). Alternatively, the yellow highlight signaled to students that there was a technical error that they needed to find and correct. He also followed a routine of writing feedback regarding students' overall performance on the task at the end of the paper using the comment function. He commented, "the students know by now, that I put that [overall feedback] on the last period there [of the paper]" (TA-12/11). Reviewing his annotations provided Mr. Sparrow with a quick way to determine *what* content or skill to focus on for that particular student. In cases where the annotations were primarily related to spelling and grammar and highlighted in yellow, the instructional decision was that the student would self-assess his/her errors and correct them. Thus, annotations facilitated and supported Mr. Sparrow's instructional decision-making process.

Mr. Sparrow also had annotation routines directed to himself. He frequently, formatively, assessed students' skill level as beginning, developing, or secure. Rather than assigning scores for this assessment he used quick annotations of 0 (beginning), /(developing), + (secure), that he noted in his Skill Development Chart (SDC). This allowed him to visually scan students' skill level. In another instance, he talked about using shorthand to make notes for himself. At mid-year, students completed a reading interest survey which Mr. Sparrow used to guide his decisions about new reading partners. As he reviewed students' responses he

commented:

So I am just going to make a note on his survey. By the way, I mark these up because, well I mark everything up, but I mark this up with just my notes, because he is not getting it back. It goes in my file so I could just do shorthand whereas if he was getting it back, I would write more elaborate notes (TA-1/20).

Mr. Sparrow circled information on students' interests using a green marker and in some places re-drafted their words or abstracted a book genre from the list of authors. This quick circling of key information and in some cases noting the genre of interest (e.g., over J.K. Rowling he wrote Fantasy) allowed him to quickly see the kinds of books students enjoyed so that he could use this information<sup>4</sup> to inform his decisions about reading partners.

#### 4.1.4. Routines for organizing student work

Organizing routines often worked with the other management routines and his interpretation strategies (discussed in the next section). Organizing routines were used for both the physical management of student work and for information inferred (e.g., grades) from that work. These routines served to put assessment events or information gleaned from them into a logical order that could be easily accessed for immediate and future interpretation.

Mr. Sparrow frequently sorted student work into piles as he made interpretations of their work. With the Post-it assignment, he reviewed each Post-it and after determining the student's skill level he physically placed the Post-it into one of three piles: beginning, developing, secure. This allowed him to visually monitor (an interpretation strategy) the skill level of the class by watching the various Post-it piles grow. Similarly, he physically sorted the students' reading surveys as he made instructional decisions regarding student partnerships. Mr. Sparrow explained:

So Wendy and Lola. I will just put them to the side and then when I align them, I just do the alternating stack kind of deal. I mean, eventually I will write them down, but I just want to get them out of my way now so I know who I am left with at the end (TA-1/20).

This physical process of "stacking the papers in pairs" with "turning the pieces of paper vertical or horizontal to differentiate one pair from the next" was a routine Mr. Sparrow used to put student work in a meaningful order (TA-1/20). In the quote above Mr. Sparrow indicated "eventually I will write them down" which foreshadowed his more formal organizing routines.

Student files, SDCs, and Google portfolios, were examples of more formal organizing routines. Because Mr. Sparrow recognized that he relied on reading his comments on students' Reading Responses and his need to see their work over time, this led him to organize each student's work in a separate file. He noted, "Right now I would say my biggest system is the file cabinet. I have a file for every student and I go through those things often" (II-10/29). This file system worked in conjunction with his Friday folder collection routine in which he would ask for some work to be returned to him after students and parents reviewed the feedback. Returned work was then preserved in students' files so it could be accessed and used for further interpretation. This routine was essential for Mr. Sparrow to track and understand student progress, because, as he stated:

The only way I can do that [check student progress] is if I have the actual responses, I do not think a rubric [score] would tell me everything I would need to know. A rubric would not help me remember everything I need to remember about that student (II-10/29).

Mr. Sparrow developed a formal routine of data organization in the form of a weekly SDC that he maintained electronically in a multi-tab spreadsheet for students individually and for the class as a whole. The chart included a list of literary skills students were expected to master over the course of each unit. For instance, in the character unit, skills included archetype, character theory, and traits, along with skills that spanned several units such as use of text evidence, prediction, and cause and effect. In this chart he tracked students' skill development as beginning, developing, or secure (using the annotations described above). By using the spreadsheet tabs he could review these data by skill, class, or student. This organization routine, like the others, provided the structure needed for Mr. Sparrow to be able to engage in interpretation strategies. Specifically, this routine insured that Mr. Sparrow collected evidence of student learning that was directly related to his learning objectives, thus providing support that his evaluations were valid.

#### 4.1.5. Routines for returning student work

Each Friday students' graded work was sent home in a "Friday folder" so that it could be shared with parents/guardians. Mr. Sparrow explained:

I do a Friday folder. I do so much of my grading outside of school that it is a way of keeping me honest and making sure that I am sending work home so parents can see it. Everything that I grade goes in the Friday folder. There are two parts to the folder: keep or send back. I keep every single one of their Reading Responses. Sometimes, I will photocopy them, and then send them home or usually what I will do is write 'send back' and I'll ask them to come back. Because that [the reading response] really shows me their growth as readers, not just their score, but the comments. If I am seeing a lot of my comments, the same comments, then that informs the conferences that I have with them here [in the classroom] (II-10/29).

Mr. Sparrow's Friday folders functioned (a) to manage his time and self-regulate ("keep him honest"), (b) as a delivery of evaluated work to students and families, and (c) as a mechanism for getting some of that work back for future use. As with the other routines described, Mr. Sparrow taught this routine to his students and also to parents/guardians so that everyone knew when (and where) to look for returned work. His comments on student work became another source of data that informed the substance of his conference discussions with students.

Mr. Sparrow also relied on his student number system to return assignments for conferencing purposes. That is, during the week (prior to sending work home) Mr. Sparrow referenced reviewed Reading Responses as content for some of his conferences. These Reading Responses were typically in numerical order to support his conferencing cycles. The routine of returning the work in numerical order facilitated his goal of conferencing alphabetically, which insured that he reached every student each week. This routine supported Smith's (2003) assertion that validity in classroom assessment is established through sufficiency of information and not traditional indices of reliability. When collecting evidence of student learning one must ask whether or not there is "enough information here to make a reasonable decision about this student

<sup>4</sup> Mr. Sparrow also used reading level scores from the beginning of the year, his most recent skill development chart, and his knowledge of the students to make these pairings.



with regard to this domain of information?” (Smith, 2003, p. 30). Mr. Sparrow consciously planned and implemented this routine to ensure that he gathered sufficient information on all of his students.

However, as with other routines Mr. Sparrow used a subroutine as part of returning student work practice. In referring to the return of reading responses for conferences he stated:

I have Lily's [Reading Response] on top, she is number 15. What I do is, the reading responses that I know I want to go into in a conference, I put those on top of the pile so it is a visual reminder of the first ones I have to go to (TA-12/5).

He used the subroutine of putting the papers of the students who he needed to meet with immediately at the top of the pile as a reminder to himself. The use of this assessment management routine supported his instructional decision of *who* to meet with first and *when*, illustrating how assessment management routines supported the implementation of instructional decisions.

Students completed longer writing assignments using Google Docs, which allowed Mr. Sparrow to provide timely feedback on their work. Mr. Sparrow described the use of Google Docs as a routine for interacting with students around their writing:

They just got their second drafts in, so the last time we talked I was saying how it is an issue if I see a student when he/she is in the pre-writing phase and then I don't see that student again until revision. Well now at least, I can type comments, students can comment back to me and then it is a much more efficient exchange. So even if time does not allow me to meet with every single student in the lesson, students can go and log on to Google Docs and see my comments (TA-12/5).

The routine of using Google Docs to collect and return student work allowed him to provide feedback more often, to stay aware of students' progress on their writing, and to be able to access and refer back to these exchanges (comments between him and the student) quite readily.

#### 4.2. Interpretation strategies

Interpretation strategies refer to the procedural knowledge Mr. Sparrow enacted to help him understand students' current level of performance on written assignments. Specifically, we identified two families of interpretation strategies: appraisal and monitoring. Through the use of these strategies (see Table 2) Mr. Sparrow assigned meaning to student work that, in some cases, had direct implications for his next instructional decision.

**Table 2**  
Interpretation strategies.

Strategies	Description	Specific Strategies and Supports
Appraisal	Strategies used to determine the quality of student work and/or the usefulness of the assessment data	<ul style="list-style-type: none"> <li>• Alignment - Objectives</li> <li>• Alignment - Instructional Intervention</li> <li>• Rubric - Formatively (keep in mind)</li> <li>• Rubric - Summatively (assign grade)</li> </ul>
Monitoring	Strategies used to consider ongoing progress (or regression) within or across individual students and the class as a whole.	<ul style="list-style-type: none"> <li>• Overall Class Performance</li> <li>• Formal Reviewing: (SDC)</li> <li>• Informal (Pile sort; Working Memory)</li> </ul>

While we describe these strategies separately, in practice Mr. Sparrow used them in concert, such that while appraising student work he also engaged in monitoring informed by the meaning he made of that work.

##### 4.2.1. Appraisal strategies

We use the term *appraisal* to describe strategies Mr. Sparrow employed when determining the value or quality of student work and when he considered the validity of the assessment task in context. We identified four strategies Mr. Sparrow used to make determinations about the quality of student work: (a) alignment of student performance with the standards taught for the assessed target skill or objective; (b) need for instructional intervention; (c) use of a rubric for formative and summative purposes; and (d) consideration of the overall class performance.

When making appraisals about a student's skill level, Mr. Sparrow considered the alignment of student work with the instructional objectives. Referring to his Post-it assignment, Mr. Sparrow elaborated:

I want to see them mention their character theory and their prediction and then have a realistic prediction based on their character theory. So that is what I'm looking for because that was the skill we went over today (TA-12/10).

We see this alignment of appraisal to instructional objectives in his interpretation of Tanya's Post-it (Fig. 1, transcribed below), which he read aloud and then reacted to:

So Tanya wrote: "*prediction (using my character theory) I do not believe my Miyax will stop at the town. She will probably keep going until she reaches Point Hope.*" See that to me is concerning because she did not make a reference to the character theory and she also left out things that we have worked on with predicting like text support, text evidence, and just backing it up. We have talked about predictions and we have talked about "skilled" predictions. I do not see any of that (TA-12/10).

As Mr. Sparrow continued to evaluate the Post-Its it was evident that he held clear standards for performance quality that were directly related to the instructional objectives he identified and taught toward. In appraising students' work, he used those same standards to guide his determination of the quality of that work. These standards were aligned with both the content of his instruction and the level of thinking he expected of his students. Thus, he attended to both test content and response process validity evidence (AERA, APA, NCME, 1999). Response process evidence addresses the nature of the response that is expected. In the example above, although the student addressed the content she did not do so in a way that aligned with the kind of cognitive processes Mr. Sparrow expected: "skilled predictions."

Mr. Sparrow's appraisals were also informed by his degree of concern for instructional intervention. Mr. Sparrow described his overall appraisal process for the Post-it assignment in the following way:

So, I'm going to separate [the Post-its] into beginning, developing, secure, or the other way I look at it is 'conference right away,' 'get to by the next conference' or 'use a different skill or strategy in the next conference' (TA-12/10).

Mr. Sparrow rated students' demonstration of the skills based on the kind of instructional decision that was warranted. Student work identified as "beginning" was associated with the determination

that those students needed an immediate conference on the topic. In contrast, students whose work was rated as “secure,” needed to be exposed to a new skill or strategy in their next conference so that they could move forward. Thus, the use of appraisals as an interpretation strategy, supported Mr. Sparrow’s instructional decision processes of determining *who*, *when*, and *what* to prioritize.

Mr. Sparrow used rubrics to support his appraisal of student work in two key ways: to guide formative feedback and to assign summative grades. Mr. Sparrow’s review of students’ work was informed by his awareness of what was on the rubric, he explained:

What I do is I have the rubric in mind when I read over these. I do not fill in the rubric as I go over it, this first time around. I will not fill in the rubric until the students come back to me and they say they have published (TA-12/11).

This strategy emerged in Mr. Sparrow’s review of Alice’s draft of a story, she titled “A Million Ways Gone.” Before even reading the writing piece, Mr. Sparrow shared that Alice “uses writing strategies consistently” and that he expected that most of his comments would be challenging her to move to the next level. Later as he read the narrative and shared his appraisal of her published work, he stated “I’m thinking, at the end of the section that I just read, she has good detail, she has what we would call ‘show not tell detail.’ These are things that are going to be on the rubric” (TA-12/11). Thus, while his comments on her draft focused on improving her writing in a variety of ways, his awareness of the rubric standards were used in determining the quality of the work.

Mr. Sparrow also used rubrics to assign summative grades to student work. When using rubrics for summative purposes he read the entire response before assigning scores on the rubric criteria. For example, after reading a student’s Reading Response, he referenced the rubric and read aloud the first two levels of the reading skill/strategy criterion. Then he explained his scoring:

I am going to put this as [circled 2] because, again, she did show one part correctly, but she misinterpreted another part. When I say ‘partial proficiency’ (level 1) I mean she really kind of missed the boat there, so this is a two (TA-12/3).

Finally, Mr. Sparrow routinely reviewed the overall class performance to make appraisals of the student data itself. For example, as Mr. Sparrow reviewed students’ Post-it responses collectively he recognized that most of the class had not demonstrated the target skill. He commented:

There’s just too many students who do not have it. I really honestly think they are close to getting it, they just didn’t show what they needed to show today. I will look at this as kind of a bump in the road more than something that is very predominant. Part of me thinks it is [the Post-it data] not as telling as it should be and part of me thinks, based on the review, I’ll get better information after I reteach (TA-12/10).

Thus, Mr. Sparrow considered the quality of the data gathered and questioned the accuracy of the inferences formed about students’ prediction skills. As a result of this process, his instruction was directly informed, but he did not use his appraisals of the individual student Post-its to record their progress in his SDC, or for other recorded purposes because he did not perceive the information to be “as telling as it should be” of students’ actual ability.

#### 4.2.2. Monitoring strategies

Mr. Sparrow also appraised student work by using a series of

monitoring strategies that allowed him to consider the ongoing process of individual students and the class as a whole. He engaged in both formal and informal monitoring. Formal monitoring occurred when Mr. Sparrow intentionally reviewed summarized records of student work, such as the SDC or students’ folders. Referring to his SDC, Mr. Sparrow stated “So it is just a good visual reminder that if I see more zeros than not, or if I see like a patch of zeros, that is something to go over” (II-10/29). Of note, Mr. Sparrow used the SDC to look both across, at each student’s individual progress, and down the skill columns to see how the whole class was progressing. This had implications for instructional decisions. Mr. Sparrow noted that if he saw a “patch of 0’s,” it signaled to him that it was content he needed to reteach. Thus, this monitoring strategy supported Mr. Sparrow in making decisions about *who* and *what* content to review.

Mr. Sparrow also informally monitored student progress by relying on his knowledge of the students, curriculum, and shared experiences in the classroom. While reviewing students’ Post-its for the skill of predicting and use of character theory, Mr. Sparrow informally tracked students’ progress on other instructional objectives. For example, when Mr. Sparrow came to a Post-it that read “I predict that Alessi’s father will still be mean to Alessi because he didn’t seem like he only disliked Alessi because of her fatness” (TA-12/10); he responded:

This is just a side note, we have been having lessons on choosing the best word, we call it “Use A Ten.” They [the words] have to be unique, specific, and have the exact connotation, and we rank them on a scale of one to ten. So, if I say happy, the challenge is for the students to find a word that is more of a ten, to go up in the scale, things like that. She likes that activity, so every day she is writing about the same character that has a weight problem, and every day she comes up with a new word for that. I would have thought she would have run out of words a while ago, but she has not, so that is a good thing (TA-12/10).

Thus, Mr. Sparrow used his knowledge of the curriculum along with his knowledge of this student as monitoring strategies to appraise this student’s work.

## 5. Discussion

A visitor to Mr. Sparrow’s classroom would see a tranquil classroom context. Students working on tasks; assignments completed and returned with apparent ease, and learning happening. To the visitor, the “Friday Folders” and the “Homework Board” seem like ubiquitous elements that can be found in most U.S. classrooms. However, our qualitative case study approach to understanding Mr. Sparrow’s practices revealed the opposite. That in fact, that visitor misses a great deal and probably fails to understand the importance of the routines and strategies working under the surface to support students’ learning experiences and Mr. Sparrow’s instructional decision making processes. Thus, one of the benefits of using a qualitative case study methodology is that it helped to reveal that Mr. Sparrow’s classroom assessment practice was iterative, fast-paced, ongoing, and pragmatic. He developed a repertoire of assessment relevant practices, which he implemented with nuanced precision. Moreover, these practices were symbiotic, such that they supported and facilitated his overall teaching practice.

### 5.1. Still waters run deep: a repertoire of assessment relevant practices

We observed Mr. Sparrow engage in five types of assessment

management routines and two families of interpretation strategies. Mr. Sparrow used these routines and strategies as he enacted various assessment practices beginning with assigning tasks through the return of appraised student work. Mr. Sparrow used assessment management routines to systematize how he assigned, collected, annotated, organized, and returned student work. Mr. Sparrow consciously and routinely employed these routines, and communicated and explicitly taught them to his students. Routines helped students form a clear understanding of what the expectations and helped Mr. Sparrow maintain a sense of order and regularity in his classroom (Nash, 2009). Mr. Sparrow also employed two families of interpretation strategies as he evaluated students' reading response and post-it assignments. These included appraisal and monitoring strategies used in concert to help him form valid interpretations of students' work and ascertain students' strengths and areas for improvement.

The management routines and interpretation strategies enacted by Mr. Sparrow mirror the best practices in classroom assessment identified in the *Classroom Assessment Standards: Sound Assessment Practices for PK–12 Teachers* (2015).<sup>5</sup> Next, we provide four examples of how Mr. Sparrow's practices aligned with the practices described in the standards.

Recall that as part of his assignment routines, Mr. Sparrow used recurring examples and models. The purpose of this was to present a consistent and shared understanding of the purpose of the assignment. Communicating to students why a task is assigned and the expectations for successful completion of the assignment, is central to their ability and willingness to complete the task (Rademacher, Schumaker, & Deshler, 1996). Standard F2: *Learning Expectations* addressed this expectation. Specifically, Mr. Sparrow provided clear learning expectations, used a method that was consistent and understood by students, and shared student exemplars as evidence of quality performance on the task. Similarly, this assessment practice provides evidence for Standard U1: *Analysis of Student Performance*, in that students were made aware of the scoring criteria that would be used to evaluate their work, and had the opportunity to hear Mr. Sparrow describe his grading processes.

Mr. Sparrow used rubrics to evaluate students' Post-It and Reading Response assignments, a practice which is discussed in Standard U1: *Analysis of Student Performance*. The rubrics included the key skills students were expected to demonstrate throughout the unit and provided detailed descriptors of the range in performance on each skill. Typically, rubrics are used to evaluate student work after it is completed, but Mr. Sparrow also used rubrics as a mechanism to help students understand the task requirements and expectations for quality performance before they began the assignment (an element of Standard F2: *Learning Expectations*). Used in this way, rubrics signaled to students the expectations or standards for task performance (Andrade, 2000; Popham, 1997). Research by Andrade (2001) and Andrade and Du (2005) found that students who received a rubric and explanation for a writing assignment earned higher scores on their essays than students who did not receive the rubric with the assignment. The authors claimed that without a rubric, students reported feeling as if they were playing a "guessing game" to figure out how they would be evaluated, whereas the rubric gave students a sense of "direction" (p.3).

Beyond rubrics, Mr. Sparrow also enacted the appraisal strategy of aligning student performance with the target skills for that unit.

To keep track of students' performance on these skills, he recorded their progress as 0 (beginning), / (developing), + (secure), on the SDC. The use of his SDC is one example of how Mr. Sparrow exhibited the assessment practices outlined in Standard U3: *Instructional Follow up*. The standard states that teachers should track student progress, and use this information to inform their teaching and student learning. Mr. Sparrow used the information in his SDC to guide instruction and provide follow-up support for students through conferences.

## 5.2. Adjusting the sails: nuance in use of assessment relevant practice

In her 1992 article, Kagan, 1992 argued that procedural routines are the "sine qua non of teaching" (p. 162). By this, she meant that mastering a host of routines was essential to teaching. She added that once acquired, teachers could employ these routines across subject areas and with different students quite uniformly. Although we appreciate and agree with the importance Kagan placed on the need to establish and use routines, we do not believe routines are neutral or that they can be applied in a vacuum without consideration of the teaching context and students. As such, we agree with Grossman (1992) who asserted that, "management is not neutral but carries within it its own implicit theories of instruction (Edelsky, Draper, & Smith, 1983)" (p. 174).

Mr. Sparrow's implementation of assessment relevant practices was nuanced. He did not mindlessly follow these practices, but rather he was sensitive to his context and re-considered their application as needed. Recall that in the beginning of the year, Mr. Sparrow used a Post-It rubric to evaluate student work, however when his purpose for giving this assignment changed to a formative one, he no longer used the rubric as part of his appraisal strategies. Mr. Sparrow's adaptability reflects what Feldman and Pentland (2003) call the performative aspect of expertise. By this, they argued that assessment-related routines, strategies, and decisions are not just a matter of knowing more, experts also know when and how to enact practices that are sensitive to their context and why a particular practice is appropriate for that situation.

Teachers' goals for student learning should inform the assessment management routines selected to support those goals. That is, learning goals and assessment management routines are reciprocally related. The goal of classroom routines should be to facilitate learning opportunities by limiting distractions and creating a structure for all to act, taking into account the assessment climate and the particular needs of students. Thus, while we suggest that the assessment relevant practices we identified in Tables 1 and 2 can be used to inform teacher education, development, and research, we also caution against teaching them as a list of practices to engage without thoughtful reflection on and understanding of the underlying purpose of each practice in the context of teaching and a teacher's level of professional experience.

## 5.3. The clownfish and the sea anemone: symbiotic nature of assessment relevant practices

Mr. Sparrow's management routines, interpretation strategies, and instructional decisions were symbiotic; they were mutually dependent on each other. Mr. Sparrow evoked management routines to support and sustain his interpretation of student work, to aid him in enacting instructional decisions, or in concert with other management routines. For example, Mr. Sparrow used a notation system of 0, /, and +, as an interpretation strategy to appraise student work. Concurrently, he also initiated a management routine (to record the skill level on his SDC), an instructional decision (meet with student who received a 0), and a second

<sup>5</sup> Please note: we describe the practices Mr. Sparrow used from the point at which he assigned the Reading Response and Post-it tasks through returning these assignments to students. We did not observe Mr. Sparrow create the task instructions for these assignments, and therefore cannot speak to some of the standards (i.e., F1: assessment purpose, F3: assessment design).



management routine (place that student's paper on the top of the pile as a visual reminder to confer with him/her first the next day).

We first noted this symbiotic relationship during data analysis when we attempted to construct visual maps of Mr. Sparrow's practices. Our maps were highly detailed, with connecting lines in all directions. This led us to two conclusions. First, the symbiotic nature of routines, strategies, and decisions formed a complex tapestry. Second, classroom assessment practices could not be separated from the other aspects of teaching such as instruction and planning. At first glance, this may seem like an obvious observation. However, the dominant approach to providing instruction or professional development in assessment is via a stand-alone course or module, in which there is little integration of assessment content with other teaching practices (DeLuca & Klinger, 2010). Alternatively, Mr. Sparrow demonstrated an understanding assessment that was seamlessly and intentionally employed with his other teaching activities. Riggan and Olah (2011) referred to this as "orchestration" of assessment, because much like a composer needs to blend the sounds of different instruments to create a piece of music, teachers also need routines, strategies, and typologies of instructional decisions to engage in effective practice (p. 4).

## 6. Limitations

Mr. Sparrow taught in a school located in a middle-class town where few students received special services or supports, and parents were active in their children's school activities. In addition, the district policy was to support teachers and defer to them as experts in their work in the classrooms. Mr. Sparrow was able to integrate his knowledge of the curriculum, students, and assessment to determine practices that would best meet his needs to use data to inform instruction. In a less privileged school context or a context where teachers have less autonomy with respect to their assessment practices, teachers may be constrained by district and school policies which limit their ability to develop their own systems in ways that Mr. Sparrow has. Thus, the strength of Mr. Sparrow as an intrinsic and instrumental case of classroom assessment practice is that he was able to generate a series of practices and methods that provide insight into the potential of teachers' assessment practices when given the freedom to pursue their own needs and the students' collective and individual needs. Additional research is needed to develop a broader understanding of the nature of teachers' assessment processes across experience levels and teaching contexts.

## 7. Implications

Findings from this study can be used to inform the fields of teacher education, policy, and assessment research. For teacher educators, the categories identified in Tables 1 and 2 can provide an overarching framework for instruction in classroom assessment processes. Kim and Klassen's (2018) investigation into teachers' responses to classroom situations illustrated stark differences in the cognitive processing of experts and novices, with novices needing assistance in understanding the underlying principles of teaching practice. The categories of assessment relevant practices we identified can be used to form an initial schema for novices' professional development. In addition, the specificity of our findings can provide teachable examples of how these practices can be employed *in situ*. We know from research in ambitious teaching that when novice teachers were taught routinized behaviors related to that practice, it reduced some of the complexity of the task, making it easier for them to enact those practices (Leinhart & Steele, 2005). Rather than forcing novice teachers by virtue of no

instruction to invent potentially ineffective strategies and routines on their own, the framework that emerged from Mr. Sparrow's practice can be a sound starting place for them to begin developing and refining their own practice.

Policy makers and assessment researchers who determine and/or inform expectations for preservice teacher certification or evaluations of teacher quality should consider how classroom assessment expertise manifested in Mr. Sparrow's practice. In our time with Mr. Sparrow, he never used measurement terms such as validity or reliability to describe or explain his practice. However, many of his practices and routines were developed to support his ability to make valid judgements about his students' knowledge, interests, and needs. Thus, as teachers' assessment knowledge becomes more central to evaluations of teacher quality, policy makers who determine teacher certification and evaluation criteria need to recognize that how teachers communicate and engage in the work of teaching, does not always reflect the terminology or specificity recommended by scholars rooted in evaluation and measurement.

## 8. Conclusion

We engaged in a study of one expert teacher's practice to identify and understand the strategies and routines he used to manage his classroom assessment practices. Mr. Sparrow employed sets of assessment management routines and interpretation strategies, and these served to systematize and guide his analysis of student work and instructional decisions. For example, Mr. Sparrow had routines for assigning, collecting, annotating, organizing, and returning student work to manage his assessment practices. Moreover, Mr. Sparrow used appraisals and monitoring as interpretation strategies to derive meaning about the quality of student work and to determine next steps in instruction. This investigation extends the extant literature by examining *how* an expert teacher engaged in assessment for his daily practice. The evidence provided here suggests that this teacher's practice was multifaceted, complex, and transformative in ways that informed teaching and learning. Our findings can be used by researchers and practitioners to understand how teachers can and should manage and use assessment information to inform teaching at the classroom level.

## CRediT authorship contribution statement

**Helenrose Fives:** Conceptualization, Data curation, Formal analysis, Writing - original draft, Supervision, Project administration, Funding acquisition. **Nicole Barnes:** Conceptualization, Data curation, Formal analysis, Writing - original draft, Supervision, Project administration, Funding acquisition.

## Acknowledgements

This work was supported by Spencer Foundation via Grant #201400152. We are sincerely thankful to Spencer Foundation for supporting our work and to the teachers, principals, and research team members who deepened our understanding of teachers' assessment practices and who made this work possible. We contributed equally to this article. We rotate order of authorship in our scholarship. All correspondence for this article should be addressed to the first author at [fivesh@montclair.edu](mailto:fivesh@montclair.edu).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.tate.2020.103063>.

## References

- Alexander, P. A. (1997). Mapping the multidimensional nature of domain learning: The interplay of cognitive, motivational, and strategic forces. In M. L. Maehr, & P. R. Pintrich (Eds.), *Advances in Motivation and Achievement* (pp. 213–250). Greenwich, CT: JAI Press.
- Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research*, 61(3), 315–343. <https://doi.org/10.2307/1170635>.
- Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, 20(1), 20–34. <https://doi.org/10.1080/0969594X.2012.736364>.
- Alm, F., & Colnerud, G. (2015). Teachers' experiences of unfair grading. *Educational Assessment*, 20(2), 132–150. <https://doi.org/10.1080/10627197.2015.1028620>.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Federation of Teachers (AFT), National Council on measurement in education (NCME), & National Education Association (NEA). (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, 9(4), 30–32. <https://doi.org/10.1111/j.1745-3992.1990.tb00391.x>.
- Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13–18.
- Andrade, H. G. (2001). The effects of instructional rubrics on learning to write. April 18 Current Issues in Education [On-line], 4(4). Available: <http://cie.ed.asu.edu/volume4/number4>.
- Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research and Evaluation*, 10(3). ISSN 1531-7714.
- Baird, J. (2010). Beliefs and practice in teacher assessment. *Assessment in Education: Principles, Policy & Practice*, 17(1), 1–5. <https://doi.org/10.1080/09695940903562682>.
- Barnes, N., & Fives, H. (2018). *Case Studies of Teacher Data Use*. New York, NY: Routledge.
- Barnes, N., Fives, H., Mabrouk Hattab, S., & SaizdelLaMora, K. (2020). Teachers' epistemic cognition in situ: Evidence from classroom assessment. *Contemporary Educational Psychology*, 60. <https://doi.org/10.1016/j.cedpsych.2020.101837>.
- Bereiter, C., & Scardamalia, M. (1993). *Surpassing ourselves: An inquiry into the nature and implications of expertise*. Chicago, IL: Open Court.
- Beswick, J., & Sloat, E. (2006). Assessment expertise: A vital component of effective early literacy instruction. *Canadian Children*, 31(1), 14–21.
- Black, P. (2012). EPMA professionals: Servants or masters? *Measurement*, 10, 33–37. <https://doi.org/10.1080/15366367.2012.677342>.
- Black, P., & William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 531.
- Bliuc, A. M., Ellis, R. A., Goodyear, P., & Hendres, D. M. (2011). Understanding student learning in context: Relationships between university students' social identity, approaches to learning, and academic performance. *European Journal of Psychology of Education*, 26(3), 417–433.
- Bogdan, R., & Biklen, S. K. (1997). *Qualitative research for education*. Boston, MA: Allyn & Bacon.
- Bogdan, R. C., & Biklen, S. K. (1998). *An introduction to theory and methods*. In *Qualitative research for education*. Boston, MA: Allyn and Bacon.
- Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, 24, 61–100. <https://doi.org/10.2307/1167267>.
- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, 7(4), 279–301. [https://doi.org/10.1207/s15324818ame0704\\_2](https://doi.org/10.1207/s15324818ame0704_2).
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational measurement: Issues and Practice*, 22(4), 13–25. <https://doi.org/10.1111/j.1745-3992.2003.tb00139.x>.
- Brookhart, S. M. (2017). *How to use grading to improve learning*. ASCD.
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., et al. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803–848.
- Brookhart, S. M., & Nitko, A. J. (2006). *Educational assessment of students* (5th ed.). Upper Saddle River, NJ: Prentice Hall.
- Brown, L. M., & Gilligan, C. (1992). *Meeting at the crossroads: Women's psychology and girls' development*. Cambridge, MA: Harvard University Press.
- Calkins, L., & Tolan, K. (2010). *A guide to the reading workshop, grades 3-5*. Portsmouth, NH: First Hand/Heinemann.
- Cizek, G. J. (2009). Reliability and validity of information about student achievement: Comparing large-scale and classroom testing contexts. *Theory Into Practice*, 48(1), 63–71.
- Coombs, A., DeLuca, C., LaPointe-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education*, 71, 134–144.
- Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory Into Practice*, 39(3), 124–130. [https://doi.org/10.1207/s15430421tip3903\\_2](https://doi.org/10.1207/s15430421tip3903_2).
- DeLuca, C., & Bellare, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education*, 64(4), 356–372. <https://doi.org/10.1177/0022487113488144>.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419–438. <https://doi.org/10.1080/0969594X.2010.516643>.
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Approaches to classroom Assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, 21(4), 248–266. <https://doi.org/10.1080/10627197.2016.1236677>.
- Earl, L. (2012). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin.
- Edelsky, C., Draper, K., & Smith, K. (1983). Hookin' em in at the start of school in a 'whole language' classroom. *Anthropology & Education Quarterly*, 14, 257–281.
- Ericsson, K. A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative task. In K. A. Ericsson, N. Charness, P. Feltovich, & R. R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 223–242). Cambridge, UK: Cambridge University Press.
- Farrell, C., & Marsh, J. (2016). Contributing conditions: A qualitative comparative analysis of teachers' instructional responses to data. *Teaching and Teacher Education*, 60, 398–412. <https://doi.org/10.1016/j.tate.2016.07.010>.
- Feiman-Nemser, S., & Parker, M. B. (1990). Making subject matter part of the conversation in learning to teach. *Journal of Teacher Education*, 41, 32–43.
- Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly*, 48(1), 94–118. <https://doi.org/10.2307/3556620>.
- Fives, H., & Barnes, N. (2017). Informed and uninformed naïve assessment constructors' strategies for item selection. *Journal of Teacher Education*, 68(1), 85–101. <https://doi.org/10.1177/0022487116668019>.
- Fives, H., Barnes, N., Dacey, C., & Gillis, A. (2016). Assessing assessment texts: Where is planning? *The Teacher Educator*, 51(1), 70–89. <https://doi.org/10.1080/08878730.2015.1107442>.
- Gareis, C. R., & Grant, L. W. (2015). Assessment literacy for teacher candidates: A focused approach. *Teacher Educators' Journal*, 4–21, 2015.
- Glaser, R., & Chi, M. T. H. (1988). Overview. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise*. Hillsdale, NJ: Erlbaum. xv–xxviii.
- Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education*, 21(6), 607–621. <https://doi.org/10.1016/j.tate.2005.05.001>.
- Grimmett, P. P., & MacKinnon, A. M. (1992). Craft knowledge and the education of teachers. *Review of Research in Education*, 18, 385–456. <https://doi.org/10.2307/1167304>.
- Grossman, P. L. (1992). Why models matter: An alternate view on professional growth in teaching. *Review of Educational Research*, 62(2), 171–179.
- Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *ECTJ*, 29. <https://doi.org/10.1007/BF02766777>.
- Harris, K. R., & Graham, S. (1992). *Helping young writers master the craft: Strategy instruction and self-regulation in the writing process*. Cambridge, MA: Brookline.
- Hounsell, D. (1997). Contrasting conceptions of essay-writing. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning* (2nd ed.). Edinburgh: Scottish Academic Press.
- Howley, M. D., Howley, A., Henning, J. E., Gilla, M. B., & Weade, G. (2013). Intersecting domains of assessment knowledge: School typologies based on interviews with secondary teachers. *Educational Assessment*, 18(1), 26–48.
- Interstate Teacher Assessment and Support Consortium. (2011). *IntASC model core teaching standards: A resource for state dialogue*. Washington, DC: Council of Chief State School Officers.
- Joint Committee for Standards on Educational Evaluation. (2015). Classroom assessment standards: Practices for PK-12 teachers. Retrieved on February 1, 2019 from <http://www.jcsee.org/the-classroom-assessment-standards-newstandards>.
- Kagan, D. M. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research*, 62(2), 129–169.
- Kim, L. E., & Klassen, R. (2018). Teachers' cognitive processing of complex school-based scenarios: Differences across experience levels. *Teacher and Teacher Education*, 73, 215–226. <https://doi.org/10.1016/j.tate.2018.04.006>.
- Kippers, W. B., Wolterinck, C. H., Schildkamp, K., Poortman, C. L., & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education*, 75, 199–213. <https://doi.org/10.1016/j.tate.2018.06.015>.
- Lampert, M. (2009). Learning teaching in, from, and for practice: What do we mean? *Journal of Teacher Education*, 61(1–2), 21–34. <https://doi.org/10.1177/0022487109347321>.
- LeCompte, M. D., Preissle, J., & Tesch, R. (1993). *Ethnography and qualitative design in educational research*. San Diego, CA: Academic Press.
- Leinhart, G., & Greeno, J. G. (1986). The cognitive skill of teaching. *Journal of Educational Psychology*, 78(2), 75–95. <https://doi.org/10.1037/0022-0663.78.2.75>.
- Leinhart, G., & Steele, M. D. (2005). Seeing the complexity of standing to the side: Instructional dialogues. *Cognition and Instruction*, 23(1), 87–163. [https://doi.org/10.1207/s1532690xci2301\\_4](https://doi.org/10.1207/s1532690xci2301_4).
- Leinhart, G., Weidman, C., & Hammond, K. M. (1987). Introduction and integration of classroom routines by expert teachers. *Curriculum Inquiry*, 17(2), 135–176. <https://doi.org/10.1080/03626784.1987.11075284>.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Lortie, D. (1975). *Schoolteacher: A sociological study*. Chicago: University of Chicago

- Press.
- MacArthur, C. A. (2012). Strategies instruction. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: Volume 3 application to learning and teaching* (pp. 379–401). Washington, DC: American Psychological Association.
- MacLellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualisations. *Teaching and Teacher Education*, 20(5), 523–535.
- Mandinach, E. B., & Gummer, E. S. (2013). A systematic view of implementing data literacy in educator preparation. *Educational Researcher*, 42, 30–37. <https://doi.org/10.3102/0013189x12459803>.
- McMillan, J. (Ed.). (2013). *SAGE handbook of research on classroom assessment*. New York: Sage.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education. Revised and expanded from "case study research in education"*. Jossey-Bass Publishers, 350 Sansome St, San Francisco, CA 94104.
- van Merriënboer, J., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist*, 38(1), 5–13. [https://doi.org/10.1207/s15326985ep3801\\_2](https://doi.org/10.1207/s15326985ep3801_2).
- Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33(1), 49–64.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis. A methods sourcebook* (3rd ed.). Los Angeles, CA: Sage.
- Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22(4), 13–25. <https://doi.org/10.1111/j.1745-3992.2003.tb00140.x>.
- Nash, R. (2009). *The active classroom: Practical strategies for involving students in the learning process*. Thousand Oaks, CA: Corwin Press.
- Newton, P. E. (2012). Clarifying the consensus definition of validity. *Measurement*, 10, 1–29. <https://doi.org/10.1080/15366367.2012.669666>.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307–332.
- Palmer, D. J., Stough, L. M., Burdinski, T. K., Jr., & Gonzales, M. (2005). Identifying teacher expertise: An examination of researchers' decision making. *Educational Psychologist*, 40(1), 13. [https://doi.org/10.1207/s15326985ep4001\\_2](https://doi.org/10.1207/s15326985ep4001_2).
- Pellegrino, J. W. (2012). Assessment of science learning: Living in interesting times. *Journal of Research in Science Teaching*, 49(6), 831–841. <https://doi.org/10.1002/tea.21032>.
- Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Popham, W. J. (1997). What's wrong-and what's right-with rubrics. *Educational Leadership*, 55, 72–75.
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46(4), 265–273.
- Pressley, M., & Afflerbach, P. (2012). *Verbal protocols of reading: The nature of constructively responsive reading*. New York, NY: Routledge.
- Rademacher, J. A., Schumaker, J. B., & Deshler, D. D. (1996). Development and validation of a classroom assignment routine for inclusive settings. *Learning Disability Quarterly*, 19(3), 163–177.
- Riggan, M., & Olah, L. (2011). Locating interim assessments within teachers' assessment practice. *Educational Assessment*, 11(1), 1–14. <https://doi.org/10.1080/10627197.2011.551085>.
- Shepard, L. A. (1989). Why we need better assessments. *Educational Leadership*, 46(7), 4–9.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15, 4–14.
- Smith, J. K. (2003). Reconsidering reliability in classroom assessment and grading. *Educational Measurement: Issues and Practice*, 22(4), 26–33.
- Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
- Stiggins, R. (1999). Evaluating classroom assessment training in teacher education. *Educational Measurement: Issues and Practice*, 18(1), 23–27. <https://doi.org/10.1111/j.1745-3992.1999.tb00004.x>.
- Ericsson, K. A., & Smith, J. (1991). Prospects and limits of the empirical study of expertise: An introduction. In K. A. Ericsson, & J. Smith (Eds.), *Toward a General Theory of Expertise* (pp. 1–38). Cambridge, England: Cambridge University Press.
- U.S. Department of Education. (2015). *Every student succeeds act: A progress report on elementary and secondary education*. Retrieved from [https://www.whitehouse.gov/sites/whitehouse.gov/files/documents/ESSA\\_Progress\\_Report.pdf](https://www.whitehouse.gov/sites/whitehouse.gov/files/documents/ESSA_Progress_Report.pdf).
- VanSledright, B. (2002). *In Search of America's past: Learning to read history in elementary school*. New York, NY: Teachers College Press.
- Villegas, A. M., & Lucas, T. (2002). Preparing culturally responsive teachers: Rethinking the curriculum. *Journal of Teacher Education*, 53(1), 20–32.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749–770. <https://doi.org/10.2307/20466661>.
- Warren Little, J. (2012). Understanding data use practice among teachers: The contribution of micro-process studies. *American Journal of Education*, 118(2), 143–166. <https://doi.org/10.1086/663271>.
- Wen, L. M., Tsai, C. C., & Chang, C. Y. (2006). Attitudes towards peer assessment: A comparison of the perspectives of pre-service and in-service teachers. *Innovations in Education & Teaching International*, 43(1), 83–92.
- Wiggins, G. (1998). *Educative assessment. Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass Publishers.
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice*, 11(1), 49–65.
- Wilson, D. (2004). The interface of school climate and school connectedness and relationships with aggression and victimization. *Journal of School Health*, 74, 293–299.
- Yin, R. K. (2017). *Case study research and applications: Design and methods*. Thousand Oaks, CA: Sage publications.