Using Performance Task Data to Improve Instruction

Amy L. Abbott & Douglas G. Wren

Published online: 29 Feb 2016.

Submit your article to this journal ⬀

View related articles ⬀

View Crossmark data ⬀

Routledge
Taylor & Francis Group

Using Performance Task Data to Improve Instruction

Amy L. Abbott and Douglas G. Wren

Department of Education, Old Dominion University, Norfolk, Virginia

ABSTRACT

Two well-accepted ideas among educators are (a) performance assessment is an effective means of assessing higher-order thinking skills and (b) data-driven instruction planning is a valuable tool for optimizing student learning. This article describes a locally developed performance task (LDPT) designed to measure critical thinking, problem solving, and written communication skills of elementary and middle school students in a large school district. An investigation was conducted to determine how middle school teachers use LDPT responses to inform instruction and to ascertain how students are reflecting on their performance. Results from an analysis of documents indicated teachers were engaged in formal data-informed practices within professional learning communities. Moreover, educators examined student work samples and planned instruction based on student-specific needs. Strategies to reinforce proficiency in the skills assessed by the district's LDPT are reported.

Introduction

Americans have become increasingly disillusioned with standardized testing brought about by No Child Left Behind (NCLB 2002). Although these state-mandated tests are used to make important decisions about students, teachers, and schools, only 26 percent of over 10,000 elementary, middle, and high school teachers surveyed for *Primary Sources 2012: America's Teachers on the Teaching Profession* (Scholastic and Bill & Melinda Gates Foundation 2012) indicated that standardized test results accurately reflect student achievement. According to the 2013 *Phi Delta Kappan/Gallup Poll of the Public's Attitudes Toward the Public Schools* (Bushaw and Lopez 2013), three fourths of Americans believe increased standardized testing in their local schools either hurt instruction or had no effect. The recent backlash against testing comes as no surprise. The *New York Times* recently reported that over 165,000 students—one in every six eligible test takers—opted out of at least one of two standardized tests mandated by New York State during the 2014–2015 school year (Harris and Fessenden 2015).

The advent of NCLB also resulted in data-driven decision making becoming standard practice in education (Messelt 2004). Principals and teachers were immersed in student, classroom, school, and district-level data and told "good data are as much a resource as staff, books, and computers" (National Forum on Education Statistics 2004, 3). Unfortunately, data-driven decision making at the school level has not had universal success, in large part due to shortcomings in the following areas: (a) building administrator preparation for using data to inform decision making, (b) teacher preparation on how to use data for instructional decision making, and (c) a clear vision or strategic plan for data-driven decision making (U.S. Department of Education 2010).

Performance assessments are making a strong comeback after nearly disappearing from the state and national scene around the turn of this century (Stecher 2010). Making a case for the increased use of performance assessments to measure higher-order thinking skills, Darling-Hammond (2014) stated, "Two decades of research has found that when teachers use, score, and discuss the results of high-quality performance assessments over time, both teaching and learning improve" (11). The terms "performance assessment" and "performance task" are sometimes used interchangeably, but most experts view them differently. A performance assessment is a compilation of performance tasks, while a performance task is defined as "a structured situation in which stimulus materials and a request for information or action are presented to an individual, who

CONTACT Amy L. Abbott ✉ alabbott@odu.edu 🖥 Teaching and Learning, Department of Education, Old Dominion University, Norfolk, VA 23529.

Table 1. Rubric elements and prompts.

Skill	Alignment	Definition
Critical Thinking	Prompt 1	Decides if information is correct and believable
Problem Solving	Prompt 2	Makes a choice and gives reasons for the choice
Written Communication	Prompt 2	Presents information and ideas that are clear, organized, detailed, and written for the intended audience

generates a response that can be rated for quality using explicit standards" (Stecher 2010, 3).

The purpose of this article is to describe how the results of a locally developed, district-wide performance task are used to make instructional decisions in middle schools. The first section provides background information and an overview of the performance task. The remainder of the article examines and explains how middle school teachers are using performance task data to inform instruction as well as how students are reflecting on their task responses and overall performance.

District performance task: Background

In 2008, the school board of a large district in Virginia adopted a strategic plan that focused on teaching and assessing specific skills students would need to thrive in the twenty-first century (e.g., effective communication and collaboration, critical and creative thinking, innovation and problem solving). To measure these twenty-first-century skills at the high school level, the district began administering a nationally normed performance task, the *College and Work Readiness Assessment* (CWRA; Council for Aid to Education 2007). Subsequently, selected district staff members have developed similar performance tasks for elementary and middle school students. Like the CWRA, each locally developed performance task (LDPT) presents an engaging, real-life scenario and related documents that students analyze before typing responses to two open-ended prompts. Student responses are scored using a four-point rubric designed to assess three skills. Table 1 shows the operational definition of the skills and their alignment with the prompts. It should be noted that responses to Prompt 2 are scored separately for problem solving and written communication.

Different versions of the LDPT have been administered twice annually to all students in grades 4 and 7 since the 2010–2011 school year. Responses to the fall task are scored primarily by teachers at each school.

The purposes of the fall task are to expose students to a low-stakes performance task and provide data for formative use by teachers and administrators at the student, classroom, and school levels. Spring task responses are scored centrally by a trained cadre of teachers; the results are used summatively. Individual scores are available to students and parents, and aggregate results are used to gauge progress on the goals of the district's strategic plan. The spring scoring process involves independent ratings by two trained scorers. An expert scorer provides a third and deciding rating if the first two scores do not match. From 2012 to 2014, the percentage of responses with exact agreement between the initial scorers ranged from 68 percent to 84 percent for the three task elements.

To obtain content validity evidence for each task, experts appraised "the relevance and representativeness of the test content in relation to the content of the behavioral or performance domain about which inferences are to be made" (Messick 1990, 5). Because evidence of content validity is not enough to demonstrate an assessment is measuring what it was designed to measure, two criterion validation studies were conducted. The criterion measure was the California Critical Thinking Skills Test (CCTST; Facione and Gittens 2012). Age-appropriate versions of the CCTST and the LDPT were administered to samples of 4th- and 7th-grade students during the spring task administration window. As shown in Table 2, all of the correlations between the task element scores and CCTST overall scores were significant at the .01 level. Values of .3 or greater suggested that the LDPT and the CCTST were measuring, to a moderate degree, the same general construct (Cohen 1988).

Investigation design

The purpose of this investigation was twofold: first, to understand better the ways middle school teachers used LDPT data to inform instruction; and second, to discover the ways in which students reflected on their responses and scores on the task. Both initiatives required taking a closer look at the processes occurring both inside and outside of the classroom setting.

Professional learning communities

Continuous school improvement requires teachers serving as learners, willing to invest in their own professional growth. In this light, the district has

Table 2. Correlations between performance task elements and CCTST overall score.

Level	n	Critical Thinking	Problem Solving	Written Communication
Grade 4	207	.36*	.33*	.35*
Grade 7	395	.37*	.31*	.42*

Note. *$p < .01$, 2-tailed.

encouraged best practices to accompany the LDPT, such as the formation of professional learning communities (PLCs), ultimately to heighten student achievement, and the employment of data-informed decision-making procedures for instructional planning and assessment (DuFour, DuFour, Eaker, and Many 2006; Hollingsworth 2012).

A district's first step in creating data-driven school cultures is to use assessment as an effective teaching tool (Stiggins 1991). Two options were offered to schools for their analysis of student responses to the fall performance task:

- Option 1 consisted of engaging staff in a school-wide effort to grade every 7th-grade-student responses to the LDPT.
- Option 2 consisted of a formal data-driven improvement planning process (DDIPP) in which staff score a sampling of 7th-grade-student assessments (Kennedy and Schumaker 2009).

Both options allowed for the identification of trends in student strengths and weaknesses. Because teachers' capacity with data analysis is connected to their instructional knowledge (Datnow, Park, and Wohlstetter 2007), creating a data-informed culture widely depends on the teacher's abilities to (a) outline clearly content standards, (b) establish scoring consensus with criterion-based rubrics, and (c) identify overall student strengths and areas in need of improvement (Kennedy and Schumaker 2009).

Schools, documents, and analysis

Naturally, the pedagogical context and circumstances at each of the nine middle schools involved in this investigation were unique. However, they did share the following criteria:

- Each school had groups of educators (i.e., classroom teachers, resource specialists, and administrators) that met regularly in PLCs to plan for instructional practice and student learning.
- Each school utilized a method to score 7th-grade student performance tasks using rubrics.

- Each school engaged in a data-driven, systematic way of analyzing student LDPT scores.
- Each middle school was diverse (e.g., socioeconomically and culturally) among student populations (i.e., gifted, special populations, English Language Learners).

There is both a place and a function for documents in qualitative research, mainly as they are examined for understanding empirical knowledge (Bowen 2009; Corbin and Strauss 2008). In this investigation, using documents as a primary data source was the most appropriate method to gain a comprehensive understanding of the teaching and learning practices in specific middle school settings.

As previously noted, the district's aim was to foster professional growth with performance assessment and data-informed practices. Central office personnel created a documentation form entitled *Plan for Analyzing, Communicating, and Using Results* (PACR) so that analysis procedures at each school were shared for accountability purposes. The form consisted of the following four sections for completion:

- Results (list your findings from your analysis)
- Implications for Instruction (identify next steps for planning and instruction)
- Communication Plan for Teachers (specify how LDPT results will be shared with teachers)
- Plan for Student Reflection (specify how students will reflect on their performance).

The PACR forms were completed through a collaborative process that involved administrators, classroom teachers, resource teachers, and school improvement specialists.

Analysis of these documents was chosen to track development of teacher practice and student reflection, first as a district-level overview to search for patterns and trends, and then at the individual school level to identify interventions and scaffolds for support. Both content and thematic analysis made up the iterative process of superficial examination, determining pertinent and nonpertinent information, and conducting a thorough examination of the documents for

final interpretations (Bowen 2009; Corbin and Strauss 2008). More specifically, general understandings were obtained, such as the option chosen by each school for LDPT data analysis. Anything not pertinent to the two research questions was eliminated from analysis.

During careful reading and coding, deeper impressions (i.e., categories) were developed from emerging patterns of codes. Categories consisting of similar characteristics were collapsed into broader themes that were defined and featured in a final codebook with examples of supporting evidence from the PACR documents (see Appendix A).

Findings

Primary understandings emerged from the document analysis with regard to the middle school use of LDPT data: (a) a data-driven process was a necessary function to enhance the use of performance tasks, (b) instructional interventions were vital components to support the skill sets assessed, and (c) data analysis with students was a vital process of self-monitoring and reflection. These predominant understandings regarding the use of LDPT data at the middle school level were developed from the supporting evidence provided in the PACR documents. They are explained in further detail in the following sections.

Data-informed instruction

We found that six of the nine schools chose Option 2, engaging in the more formal protocols of the DDIPP. During this formal process, schools reported engagement in the following procedures: (1) randomly selecting 16 or more student responses, (2) discussing and documenting trends in student responses—particularly areas in need of improvement, and (3) planning the next instructional steps in the classroom, such as flexible grouping of students for differentiated instruction. These results were included in a comprehensive report and shared with teachers at each of the six schools. The three other schools chose Option 1, or scoring every 7th-grade LDPT response, after the entire staff was trained on the use of the rubric. After scoring, student scores were compiled in Excel spreadsheets, and these data were shared with 7th-grade teachers.

Commonly reported among the nine schools was the collaboration of staff members in PLCs to look at LDPT data and determine the next steps needed in the classroom. Specifically, the development and enactment of data-driven processes included the support of various specialists such as gifted, special education, and reading resource teachers. Their role as leaders in data-analysis processes was instrumental while using student LDPT results to plan with teachers. Moreover, these specialists served as key instructional leadership team members, charged with generating action plans and professional development based on the identified needs of the school.

Instructional interventions

We concluded that the instructional interventions reported comprised various strategies, stretching across all core subject areas, and in some cases art, music, and health classes. Examples of strategies that reinforced critical thinking in the classroom included students engaging in Paul's Reasoning Model to support arguments (Paul 1993), partaking in evaluative scientific labs, analyzing political cartoons, and employing Socratic seminars. As for problem-solving strategies, examples included teacher-generated problem-based learning tasks, interdisciplinary performance tasks, and DBQs (document-based questions) in history. To aid in written communication—not just in English class but across all subject areas—classroom interventions included teacher-generated writing prompts, a boost in persuasive writing techniques, and the incorporation of RAFTS (role, audience, format, and topic). As such, the skills assessed in the LDPT (i.e., critical thinking, problem solving, and written communication) were incorporated into daily classroom practice. These strategies were implemented to place student learning as the primary focus.

Student reflections

The nine middle schools reported several different methods of sharing LDPT scores with their respective students (see Appendix B). At most schools, teachers modeled sample responses representing each of the four levels on the scoring rubric (i.e., novice, emerging, proficient, and advanced). At a few schools, students worked collaboratively and used the sample responses to interpret their scores in accordance with the rubric. Moreover, other schools returned copies of students' scored responses with accompanying rubrics for comparison.

Table 3. Student plus/delta responses.

Plus (positive)	Delta (needs improvement)
Critical Thinking (CT) • I used statistics from the supporting documents to make my point; I provided facts and no opinions. • My requests for missing information were logical. • I gave many examples of unbelievable information.	Critical Thinking (CT) • I could have included more information that was misleading or not believable and explain why. • I needed to include more facts and reasoning from the documents. • I should have included more questions about what I needed to know more about.
Written Communication (WC) • I had good transition words and addressed the right audience with clear information. • My response was long enough and detailed to say what I needed to say. • I stated my main points with strong responses and evidence.	Written Communication (WC) • I had too many run-on sentences. • I should have checked my spelling and grammar for errors. • I could have used proper opening and closing sentences.
Problem Solving (PS) • I recommended one of the choices and supported my choice with two or more reasons. • I used information from the passages in my answer. • I stated my answer firmly in the beginning of the paragraph.	Problem Solving (PS) • I did not use all of my resources in my answer. • I should have chosen just one recommendation and made it clear. • I did not include four distinct reasons for my answer.

As a result of analyzing the fall responses, both teachers and students engaged in reflective practices. For example, an *Instructional Interventions* form (see Appendix C) was used by teachers at one middle school to promote conversations during PLC meetings and plan for instructional interventions in the classroom. To obtain feedback from students, a *Plus/Delta* form (see Appendix D) was used at one middle school to capture perceived areas of student strength and areas in need of improvement.

Shown in Table 3 below are student responses collected from the *Plus/Delta* student self-monitoring and reflection form. Pluses were referenced as strengths of the students' work, while deltas referred to the areas students felt needed improvement. Ultimately, the opportunity to reflect on their work allowed students to think about their upcoming performance on the spring LDPT.

Four of the middle schools reported their 7th-grade-students' engagement in the best practice of goal setting, to which there is an element of metacognition. Particularly, such engagement is associated with self-regulated learning, to be accompanied by strategic planning during task analysis (Zimmerman 2008). The teacher actions of showing students how to self-regulate learning through experience, reflection, and social interaction demonstrated the formative nature of education with students.

Discussion

The PACR has served as a purposeful document that has contributed to the understanding of protocols and procedures taking place across the district with regard to fall LDPT assessments. For example, we learned the majority of the nine schools used data analysis protocols, such as the DDIPP, to examine student performance trends related to twenty-first-century skills. Their engagement in professional learning communities allowed for the analysis of LDPT data as a part of planning for continuous school improvement. Specifically, teachers used student scores to make informed pedagogical decisions, modify instruction for individuals and small groups, identify target areas within the teaching and learning process, and refine instructional tasks to support student performance on the LDPT skill sets. Ultimately, the performance task described in this study has helped change the focus of classroom instruction and assessment in many schools.

The actions of the district to develop and enact the LDPT supported the call for transformative assessment practices for successful educational reform, to which an essential component is the use of formative assessment (Black and Wiliam 1998; Gordon Commission on the Future of Assessment in Education 2013; Rabinowitz 2010; Stiggins 2008). Formative assessment is more effective when learners are given opportunities for self-reflection and peer-reflection (Black and Wiliam 1998). Students can capitalize on performance assessment experiences when constructive feedback is received from teachers, peers, or both, and by establishing learning goals to assume personal ownership of their educational domains. Although not all students in our investigation engaged in self-reflection, we consider this an opportunity for future reform.

Conclusion

Using data from a district-wide assessment to evaluate student proficiency required the collective efforts

of key stakeholders—teachers, school administrators, and central office personnel (Wayman, Jimerson, and Cho 2012). The main limitation of this investigation was the lack of follow-up to determine whether the data analysis procedures or student reflections were effective or had any impact on learning or on subsequent performance task achievement. Moving forward, we recognize the need for a reporting measure, or follow-up document, to be implemented at each school to ascertain if and to what extent the instructional interventions are impacting student learning.

Based on the outcome of our investigation, we believe performance assessment has the potential to change teaching on a scale that curricular revisions and typical professional development activities cannot achieve. "In other words, the use of the instruments should cause teachers to teach in such a way as to foster critical thinking in their various subjects" (Paul and Elder 2007, 7). While the performance task described in this investigation was exclusive to the district, the ideas presented may be useful for administrators and teachers in their work with performance assessment and data-driven instruction.

References

Black, P., and D. Wiliam. 1998. Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan* 80 (2): 139–48.

Bowen, G. A. 2009. Document analysis as a qualitative research method. *Qualitative Research Journal* 9 (2): 27–40.

Bushaw, W., and S. Lopez. 2013. The 45th annual PDK/Gallup poll of the public's attitudes toward the public schools. *Phi Delta Kappan* 95: 9–25.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Corbin, J., and A. Strauss. 2008. *Basics of qualitative research: Techniques and procedures for developing grounded theory* 3rd ed. Thousand Oaks, CA: Sage.

Council for Aid to Education. 2007. *College and work readiness assessment*. New York: Council for Aid to Education.

Darling-Hammond, L. 2014. Testing to, and beyond the Common Core. *Principal* 93 (3): 8–12.

Datnow, A., V. Park, and P. Wohlstetter. 2007. *Achieving with data: How high-performing school systems use data to improve instruction for elementary students*. Los Angeles, CA: USC Center on Educational Governance.

DuFour, R., Re. DuFour, R. Eaker, and T. Many. 2006. *Learning by doing*. Bloomington, IN: Solution Tree Press.

Facione, P. A., and C. A. Gittens. 2012. *California critical thinking skills test: M series test manual*. Millbrae, CA: Insight Assessment/California Academic Press.

Gordon Commission on the Future of Assessment in Education. 2013. *A public policy statement*. Princeton, NJ: The Gordon Commission. http://www.gordoncommission.org/rsc/pdfs/gordon_commission_public_policy_report.pdf.

Harris, E., and F. Fessenden. 2015. "Opt out" becomes anti-test rallying cry in New York state. *The New York Times*. http://www.nytimes.com/2015/05/21/nyregion/opt-out-movement-against-common-core-testing-grows-in-new-york-state.html?_r=0.

Hollingsworth, L. 2012. Why leadership matters: Empowering teachers to implement formative assessment. *Journal of Educational Administration* 50 (3): 365–79.

Kennedy, B. D., and M. R. Schumaker. 2009. *Data-driven improvement planning process* (Unpublished instrument). http://www.vbschools.com/online_pubs/kaleid/2011/02/index.asp.

Merriam, S. B. 1998. *Qualitative research and case study applications in education: Revised and expanded from case study research in education* 2nd ed. San Francisco, CA: Jossey-Bass.

Messick, S. 1989. Validity. In *Educational measurement* 3rd ed., ed. R. L. Linn, 13–103. New York: Macmillan.

National Forum on Education Statistics. 2004. *Forum guide to building a culture of quality data: A school and district resource* (NFES 2005–801). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425. 2002. http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf.

Paul, R. 1993. *Critical thinking: What every person needs to survive in a rapidly changing world* 3rd ed. Dillon Beach, CA: Foundation for Critical Thinking.

Paul, R., and L. Elder. 2007. *Consequential validity: Using assessment to drive instruction* (White paper). Tomales, CA: Foundation for Critical Thinking. http://www.criticalthinking.org/files/White%20PaperAssessmentSept2007.pdf.

Rabinowitz, S. N. 2010. Next-generation assessment systems. *Education Week* 29 (22): 36.

Scholastic and the Bill & Melinda Gates Foundation. 2012. *Primary sources 2012: America's teacher on the teaching profession*. New York: Scholastic. http://www.scholastic.com/primarysources/pdfs/Gates2012_full.pdf.

Stecher, B. 2010. *Performance assessment in an era of standards-based educational accountability*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

Stiggins, R. J. 1991. Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice* 10 (1): 7–12.

———. 2008. *Assessment manifesto: A call for the development of balanced assessment systems*. Portland, OR: Educational Testing Service, Assessment Training Institute.

U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. 2010. *Use of education data at the*

local level: From accountability to instructional improvement. Washington, DC: USDOE. http://www.ed.gov/about/offices/list/opepd/ppss/reports.html#edtech.

Wayman, J. C., J. B. Jimerson, and V. Cho. 2012. Organizational considerations in establishing the data-Informed district. *School Effectiveness and School Improvement: An Interna-tional Journal of Research, Policy and Practice* 23 (2): 159–78.

Zimmerman, B. J. 2008. Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal* 45 (1): 166–83.

Appendix A

Performance Assessment Data and Results

Theme	Category	Strategies and Techniques
Data-Driven Process Defined as a necessary function to accompany performance tasks (i.e., IPT)	Professional Learning Communities	Groupings by grade/subject level, departments (e.g., English), staff professional development sessions, Instructional Leadership Teams, specialists (e.g., Gifted, Reading Resource, Math, SPED)
	Data-Informed Protocols	Analyze data on spreadsheets, weaknesses and strengths, instructional planning, analyze trends, flexible grouping of students
Instructional Interventions Defined as vital components to support the skill sets assessed in the performance task (i.e., IPT)	Written Communication (WC)	Self-editing tools, rubric evaluation, bottom-up editing, ARMS (i.e., add, remove, move, substitute) revision, teacher-generated prompts, graphic organizers, proof for/proof against, restate written rules, rubric construction, journaling, best practices for SOL Writing, persuasive writing techniques, RAFT (i.e., role, audience, format, topic)
	Critical Thinking (CT)	Gather evidence, restate prompts, defend word choice, cite specific examples, analyze critical cartoons, support arguments, lab self-evaluations, Socratic Seminars, analyzing supporting documents
	Problem Solving (PS)	Practice IPT prompts, evaluation labs, common assessments by subjects, DBQs (i.e., document-based questions) in history, PBLs (i.e., problem-based learning), performance task GRASPS (goal, role, audience, situation, product/purpose, situation for success) strategy

Appendix B

Data Sharing

Theme	Category	Strategies and Techniques
Data Analysis with Students Defined as processes of self-monitoring and reflection from feedback (i.e., IPT scores)	Feedback	Teacher modeling of L1-4 responses, peer review, peer editing, "think-alouds," evaluate samples, goal-setting, self-reflection
	Sharing Methods	Provide student responses with rubrics, provide IPT scores, plus/delta format

Appendix C

Instructional Interventions Form

Part I - Use the following prompts to promote conversations during your Professional Learning Community (PLC) meeting. Select one member to take notes. Submit this form to your designated administrator upon completion.

Subject: _____

- What instructional strategies do we use that require analytical and critical thinking?
- Which of the following could we incorporate into our classroom instruction in order to improve student abilities with critical thinking, problem solving, and effective communication? Place a check mark next to any strategy or strategies you will incorporate into your instruction this next quarter.

_____ Performance Task (w/ an accompanying rubric)

_____ Socratic Seminar or Philosophical Chairs

_____ Reasoning and Justification through Written Communication

_____Application of Content Standards with a Real World Scenario

_____Other _____

How do you plan to incorporate this strategy into your instruction?

_____

_____

_____

Part II - Reflect on the results after implementing the selected strategy or strategies from above.

Appendix D

Student Reflections Form

After viewing each of your written responses, complete the following Plus/Delta. Anything you did well should be recorded on the "Plus" (+) side. Record how you should have responded differently on the "Delta" (Δ) side.

(Strand) _____

Plus (+)	Delta (Δ)