What is PDE?

Research Report

Paul Nichols

December 2013

PEARSON

About Pearson
Everything we do at Pearson grows out of a clear mission: to help people make progress in their lives through personalized and connected learning solutions that are accessible, affordable, and that achieve results, focusing on college-and-career readiness, digital learning, educator effectiveness, and research for innovation and efficacy. Through our experience and expertise, investment in pioneering technologies, and promotion of collaboration throughout the education landscape, we continue to set the standard for leadership in education. For more information about Pearson, visit http://www.pearson.com/.

About Pearson's Research Reports
Our network **mission** is to spark innovation and create, connect, and communicate research and development that drives more effective learning. Our **vision** is students and educators learning in new ways so they can progress faster in a digital world. Pearson's research papers share our experts' perspectives with educators, researchers, policy makers and other stakeholders. Pearson's research publications may be obtained at: http://researchnetwork.pearson.com/.

Abstract

This paper provides an overview of PDE for assessment design. First, some goals that

motivated the creation of PDE are presented followed by a summary of the conceptual

framework behind PDE. Activities involved in PDE will be described next, including an example

of the generative and reusable tools available as part of PDE. Finally, the role of the PDE leader

in successful assessment design and development is offered. This research report is intended for

a non-technical audience. A more technical description of PDE will be published in the future.

Anyone interested in employing PDE and learning more about PDE should contact Paul Nichols

of the Center for NextGen Learning & Assessment at paul.nichols@pearson.com.

Keywords: principled design for efficacy, assessment design

What is PDE?

Educators need assessments that evaluate higher-level skills, guide instructional decisions, and engage students' interest. Performance assessments meet this requirement by using a variety of approaches, from short, constructed-response and technology-enhanced items, to online games and simulated environments. For a description of the full range of performance assessments see the *Framework of Approaches to Performance Assessment*. The need to evaluate higher-level skills, guide instructional decisions, and engage students' interest pushes assessment designers and developers beyond traditional designs and development processes to produce quality content for all types of assessment contexts quickly and cost effectively.

Principled Design for Efficacy (PDE) is an approach intended to improve the design and development of assessment approaches, ranging from those that are more familiar (e.g., short constructed-response items and essay prompts), to more complex and extended approaches (e.g., projects, portfolios, and online games). As discussed in *The Learning Diamond: A Systemic Perspective on Student Learning*, assessment does not stand alone; it is used in a larger system of curriculum, instruction, professional learning, and the school ecosystem in which teachers teach and students learn. While current implementation of PDE is limited to the design and development of assessments, future work will more tightly connect the design and development of assessment to the Learning Diamond.

This paper provides an overview of PDE for assessment design. First, some goals that motivated the creation of PDE are presented followed by a summary of the conceptual framework behind PDE. Activities involved in PDE will be described next, including an example of the generative and reusable tools available as part of PDE. Finally, the role of the PDE leader in successful assessment design and development is offered. This research report is intended for

a non-technical audience. A more technical description of PDE will be published in the future. Anyone interested in employing PDE and learning more about PDE should contact Paul Nichols of the Center for NextGen Learning & Assessment at paul.nichols@pearson.com.

PDE Goals

PDE was created with a number of goals in mind. In this paper, we talk about three of those goals that we try to achieve with every project: (a) facilitate development of a range of assessment approaches, from traditional approaches, such as short constructed-response items and essays, to more complex, less familiar approaches, such as games and simulated environments; (b) make assessment content, e.g., items and tasks, development more efficient and effective; and (c) collect efficacy evidence for PDE, both as PDE continues to be developed as an approach to design and development, and as PDE is implemented with specific projects.

Development of More Assessment Types

The first goal in creating PDE was to support assessment content developers in the development of a range- of assessment approaches, from traditional approaches to emerging approaches. The Conceptual Framework and activities created as part of PDE are intended to help assessment content developers, who understand how students think and learn, and PDE leaders, who understand the PDE framework, think about and work through design and development decisions for all assessment approaches.

Perhaps the greatest potential for PDE is to support assessment content developers as they stretch beyond their experience at developing more conventional assessment types to creating relatively new assessment types, like games and simulations. Not only are these emerging assessment types more expensive and time consuming to develop, the measurement opportunities are often less discrete, although potentially far richer than in traditional assessment types. These

factors make the design and development of approaches, such as games and simulations, a more complex challenge.

PDE provides a framework and set of activities that, facilitated by PDE leaders, will help customers and assessment content developers explore alternative and innovative design and development decisions. These conversations should lead to decisions and shared understandings that, along with empirical evidence, will support the design and development process and achievement of the intended outcomes. Falling into the routine of conventional test design and development practices, however comfortable, may make the successful development of newer assessment types more difficult to achieve. Conventional test design and development practices evolved over decades to efficiently create large numbers of relatively inexpensive item types, such as multiple-choice items for customers who provided explicit guidance. While these guidelines, rules-of-thumb, and statistical methods produce tests that tend to satisfy technical requirements for large-scale assessments, conventional practices were not developed for creating newer assessment approaches, such as those embedded in games and simulations.

Efficient Design and Development of Effective Assessments

The second goal in creating PDE is to make question development more efficient by improving the process used to create effective assessment situations, such as items, tasks and simulations that more consistently and accurately hit what we are trying to assess. PDE makes stimuli and question development more efficient by incorporating generative and reusable tools for test development. PDE is generative because the test development tools, design guidelines, and stimuli and question templates may be used to generate additional stimuli and questions that retain the same link to content standards as the original stimuli and questions. PDE is reusable because the test development tools may be used multiple times, in multiple assessments, and

retain the thought and problem solving invested in developing the original question. Efficiencies in developing traditional multiple-choice items have been created and refined over many decades; however, with relatively more complex or new assessment approaches, such as performance tasks or simulations, reconceptualizing processes and creating generative and reusable tools is key to achieving efficiencies.

Collection of Efficacy Evidence

The third goal with PDE is to collect evidence for the efficacy of PDE, both in the development of PDE (Does PDE as a framework and set of activities meet the intended outcome?) and its implementation (By using PDE, are assessment situations developed that meet the intended outcomes?). Efficacy has been defined by Pearson as evidence of a measurable impact on student learning. Evidence must be collected demonstrating that PDE has a measurable impact on content development by, for example, resulting in higher quality questions and passages, and eventually on student learning. This occurs by building opportunities to collect efficacy evidence into PDE activities. In addition, the same evidence used to show that PDE has a measurable impact on content development, and eventually on student learning, can be used as evidence to support the validity of the intended interpretations and uses of assessment results. Finally, while the foundation of PDE has been established, its implementation on projects is limited. An expansion of the current framework and activities, the development of new reusable tools, and refinements based on lessons learned along the way using the efficacy evidence we collect as our guide is anticipated. To achieve these three goals, PDE includes a conceptual framework, a set of activities, and a number of reusable tools.

Conceptual Framework

The PDE conceptual framework describes the ideas underlying the activities and generative and reusable tools in PDE. The purpose of the conceptual framework is to support assessment professionals as they stretch beyond previous experiences and routines. .To elaborate on the concepts of construct, content, and evidence, use the assessment of students' understanding of the GED science standard, "Evaluate whether a conclusion or theory is supported or challenged by particular data or evidence" as an example" (GED Testing Service, 2013). This standard asks how well students can make a scientific argument supporting or challenging a conclusion or theory.

Construct

Construct is the concept of what is assessed, such as stages in a learning progression, a set of misconceptions, or practices in a discipline. PDE leaders and the content development teams they work with should have a clear understanding of what they are trying to assess, which means doing some homework. Typically, standards documents are policy documents and are not constructed to serve as detailed guides for content development. For many years, researchers within specific domains have studied the different types of knowledge, skills, and abilities (KSAs) represented by standards. By connecting the standards to their underlying KSAs, content development teams can take advantage of all the research about how students think and learn in relation to the standards.

As part of PDE, PDE leaders and content development teams dig into the published research—that is, do the homework—and summarize this research in three areas:

- The models and theories of the KSAs we are trying to assess and how students differ in these KSAs;

- The features of questions and scenarios researchers have used to elicit the KSAs we are trying to assess; and,

- The features of student behaviors researchers have accepted as evidence of the KSAs we are trying to assess.

Both PDE leaders and content development teams should have a clear and shared understanding of the KSAs, features of questions and scenarios, and features of student performances, so this understanding can be leveraged in assessment design and development. PDE leaders will work with content development teams to identify, discuss, and agree upon the KSAs and features. The use of published research findings in these ways sets PDE apart from other approaches to content development.

For example, researchers have studied for many years how students learn, create and use arguments in science (Osborne & Patterson, 2011; Berland & Reiser, 2009). This research has described the KSAs students use in making a scientific argument, including the following:

- Knowledge of how to make an assertion or claim about something;

- Skill in selecting and using data to support a claim;

- Knowledge of a warrant, i.e., a general belief, principle or lesson that supports using data for a claim. For example, the general belief that the co-occurrence of two phenomena suggests that one causes the other;

- Skill in using rebuttals to challenge the authority of the warrant. For example, "It is too early to draw conclusions. There haven't been any studies."

Content

For content, PDE leaders and the content development teams should understand the content of questions and scenarios as something they can manipulate and control. The same researchers who have studied KSAs also study the kind of questions and scenarios that encourage students to use those KSAs. By reviewing this research, PDE leaders and content development teams will be able to use of these findings to create more effective assessments.

For example, researchers studying how students make scientific arguments (Kind, Kind, Hofstein & Wilson 2011) have found that the questions that tend to elicit students use of argument ask students to *recognize or create* one or more of the following:

- A claim;

- Evidence, either less or more relevant;

- Reasoning that describes how coherence with the predictions of a theory and the data provided in a table is relevant to the claim; and,

- Rebuttal to the claim.

Evidence

For evidence, PDE leaders and content development teams should break from the routine of thinking about answers as right or wrong or creating rubrics to score answers between, for example, zero and four. PDE leaders and content development teams should give more thought to how students' responses are intended to be interpreted and if those interpretations fit the intended uses. PDE leaders and content development teams should ask the following types of questions:

- What kind of insights about KSAs—for example, about a particular stage in a learning progression or understanding of a science practice—should we gain from students' answers?

- What are the different types of answers that would give us insight into students' KSAs?

- How are the different types of answers mapped back to what they tell us about students' KSAs?

Again, PDE leaders working with content development teams must do their homework because the same researchers who have studied KSAs and questions and scenarios have also studied the kinds of student behaviors—e.g., answers and performances—that offer insight into those KSAs. As the research is reviewed, PDE leaders and content development teams should identify and document the types of behaviors that have been used as evidence of the KSAs we are trying to assess.

Researchers studying how students make scientific arguments (Sampson & Clark 2008) have found that students differ in the kinds of answers they give. For example, students differed in the number of argument components they could create or recognize. The easiest thing for

students to do was only to recognize a claim. The moderately-hard thing for students to do was to

create or recognize a claim and evidence or reasoning. The hardest thing for students to do was

to create or recognize a claim, evidence, reasoning, and a rebuttal.

Process Activities

In addition to the conceptual framework, PDE includes a set of activities, which help

translate the complexity of PDE into a more concrete and practical process. PDE leaders and the

content development team should think carefully and clearly, document their decisions, and then

act on those decisions. These activities are described as a guide for designing and developing

assessments. Each design and development activity builds on earlier design and development

activities.

In addition, each design and development activity includes collecting evidence that may

support efficacy and validity. In Table 1, different categories of activities and the types of

evidence collected during those activities are shown. Validity is a judgment of the degree to

which evidence and arguments support the interpretations and uses of test scores (Kane, 2006).

The same evidence used to show that PDE has a measurable impact on content development and

eventually on student learning can be used as evidence to support the validity of the intended

interpretations and uses of assessment results. If an activity is skipped, then the evidence

collected during that activity is lost.

Table 1.

Different types of evidence collected during different categories of activities

Activities	Evidence
Identify general KSAs, content features, and relevant behaviors.	Research, reviews, and questionnaires
Create stimuli and questions.	Questionnaires, alignment and depth of knowledge studies, and field test results
Construct specific cognitive models and task and rubric templates.	Questionnaires and expert reviews
Apply specific task and rubric templates to generate new stimuli, questions, and rubrics.	Questionnaires, alignment and card sort studies, and field test results

For example, PDE leaders and the content development team may engage in the design and development activity *Search the research literature to describe the kinds of questions, scenarios, and contexts that encourage students to use targeted KSAs*. The published research results they find would provide the efficacy evidence for this activity and also lend support for the eventual interpretations and uses of test scores. Next, building on the last activity, the PDE leader can complete the design and development activity *Train assessment content developers on how to manipulate and control the type of questions, scenarios, and contexts to influence students' use of targeted KSAs*. At the end of this activity, the PDE leader should administer a questionnaire asking assessment content developers if they understood the type of questions and scenarios that influence students' use of targeted KSAs and if they were confident they could use them to influence students' use of targeted KSAs. For some projects, a few activities have been

completed before starting the PDE process. In these cases, while the PDE approach was not

followed, the decisions or outcomes of the activity should still be documented so that it

contributes to the collection of evidence.

Reusable Tools

One goal in creating PDE was to make assessment content development more efficient by making reusable tools that are constructed when PDE is used. Tools are reusable because they may be used multiple times, in multiple assessments, and they retain the thought- and problem-solving invested in developing the original question, context, or some other type of content.

An example of a reusable tool is the content guidelines that are constructed when PDE is being implemented. Content guidelines identify for the assessment content developer the types of features, such as tables, graphs, ideas, or procedures, to include in questions, scenarios, or other types of content so that the content is more likely to assess what we are trying to assess. Using results of empirical research (Berland & Reiser, 2009; Osborne & Patterson, 2011) and input from science experts, we constructed a set of content guidelines for writing passages for the GED science standard "Evaluate whether a conclusion or theory is supported or challenged by particular data or evidence." These content guidelines may be reused any time we create passages for questions or scenarios that are focused on assessing how well students use scientific arguments.

Using the content guidelines for this standard, all passages should include the following:

- Present a single theory as a suitable explanation for data;

- The theory is claimed to explain how the data is caused by some behavior of an organism or system as controlled or organized by a *single, unobserved* cause, such as a region of the brain controlling animals' internal clock;

- Present evidence, such as tables, charts, or graphs of data compatible with the theory or hypothesis as a suitable explanation for the data.

When creating passages or scenarios that support asking more challenging questions, passage should include the following:

- Include **two sources** of evidence, such as tables, charts, or graphs;

- The **first source of evidence** allows a *causal inference* about the single, unobserved cause because the data were collected from an experiment deliberately varying or manipulating a treatment; and

- The **second source of evidence** is consistent with the theory, but does not support a causal inference, because the data show only a co-occurrence of two phenomena.

When creating a passage that supports asking only moderately challenging questions, the following content should be included:

- Include only **one source** of evidence, such as a table, chart, or graph;

- This **source of evidence** is consistent with the theory, but does not support a causal inference because the data show only a co-occurrence of two phenomena.

After training, passage writers indicated on a questionnaire that the content guidelines for writing passages were moderately to completely clear and these guidelines were applied when writing passages.

PDE Leader

Effective implementation of PDE requires extensive knowledge of assessment, a deep understanding of the PDE Conceptual Framework, and activities and experience using PDE for different projects with different purposes and constraints. It is powerful when used effectively but if not, it could lead to confusion, rework, and missed outcomes.

The responsibility of the PDE leader is to support the effective implementation of PDE

based on the needs and goals of a particular project and with fidelity. Although PDE is adaptable,

it must still meet the goals described earlier for PDE:

1) Facilitate thoughtful development of assessment;

2) Make item and task development more efficient and effective; and

3) Collect efficacy and validity evidence.

Next Steps

The Center for NextGen Learning & Assessment continues to develop, document, and research PDE. The Center continues to build evidence for the efficacy of PDE project by project. The greater the number and variety of projects in which PDE is successful, the higher will be the confidence level in the efficacy of PDE.

This paper is a brief introduction to PDE. For more information on PDE, contact Paul Nichols of the Center for NextGen Learning & Assessment at paul.nichols@pearson.com.

References

Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education*, *93*, 26–55. doi: 10.1002/sec20286

GED Testing Service (2013). *Assessment guide for educators*. Washington, DC: GED Testing Service.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 17-64). Washington, DC: The National Council on Measurement in Education & the American Council on Education.

Kind, M P., Kind, V., Hofstein, A., & Wilson, J. (2011). Peer argumentation in the school science laboratory-Exploring effects task features, *International Journal of Science Education, (I-first)* 33(18): 2527-2558.

Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education*, *95*, 627–638. doi: 10.1002/sec20438.

Sampson, V. & Clark, D. B. (2009). The impact of collaboration on the outcomes of argumentation. *Science Education,* 93 (3), 448-484.