Assessing Conceptual Understanding in College-Level Inquiry Labs

Lisa Hunter, Barry Kluger-Bell, Scott Seagroves,Anne Metevier, Austin Barnes, Andrew Norton, and Tiffani Quan

We acknowledge the work of many past PDP collaborators.

I. Overview

This document serves as background reading for participants in the Professional Development Program (PDP), offered by the Institute for Scientist & Engineer Educators.

Assessment is a process of inference – much like science. Educators (like experimenters) must set up situations in which relevant data can be gathered about learners; in the light of a model of how we *think* learning progresses, these data provide the evidence on which the educator bases a judgment of student progress. However, before any data can be gathered educators must know very clearly the goal and what the intended learning outcomes are. The PDP reading (Wiggins & McTighe, 1998) emphasizes the "backward" design process, starting from learning goals, defining acceptable evidence that learners have reached those goals, how that evidence will be elicited, and then designing instruction. We will use backward design to guide us throughout the PDP, including the integration of a culminating assessment task into all PDP-designed activities. Participants will gather evidence through a carefully crafted prompt in their culminating assessment task, and will make a judgment on learners' progress using an assessment tool called a rubric. By the end of the PDP we want participants to evaluate their activity design and teaching in relation to intended learning goals, using evidence gathered formally in the culminating assessment task and informally throughout their teaching. A key part of assessment in the PDP is to disentangle content and practice, so that each can be taught and learned in a way that can be applied to new contexts (see Box 1). This document outlines the process participants will engage in throughout the PDP, but in summary we expect that participants will evaluate their design and teaching by considering:

- Assessment of learners' understanding of the main content learning goal, based on evidence gathered formally in a culminating assessment task, and informally during teaching
- Observations of learners' engagement with the main STEM practice
- Whether the assessment designed provides the evidence needed to evaluate learners in relation to the most important learning goals

Box 1: Disentangling intertwined content and practice

As described in the PDP Inquiry Theme, content and STEM practices are intertwined. They might even seem to be so intertwined that they are impossible to disentangle. A scientist comes up with a question or hypothesis about an area of content, and as he or she conducts investigations there can be revisions to the question. An engineer designs not by trial and error, but guided by scientific principles. Content motivates and drives STEM practices, and STEM practices are used to gain new understandings or design new tools. However, in order to teach, learn, and assess it is extremely helpful to disentangle content and practice. That is, a clear picture of what it looks and sounds like when a learner understands a concept, for example with a few key things the instructor is looking for, makes teaching, learning and assessing much more feasible. And likewise, having a clear understanding of the generalizable aspects of a STEM practice, that make it transferrable to new contexts, makes it something that can be taught, learned and assessed. One might consider this approach similar to the scientific/engineering process of finding ways to identify and isolate aspects of complex systems, to better understand them.

It is the reflective process that we value most, and expect that by the end of the PDP experience you will have practiced in a very concrete way the backward design process, and used it to evaluate your design and teaching. We hope very much that participants embrace the opportunity to design, teach, and assess learners for their own professional development, instead of to assign a grade. Assigning grades and evaluating performances in other ways are a reality, and getting experience with assessment tools and strategies is a key part of this. By carefully performing a very focused and concrete assessment activity, we expect that participants gain new tools and perspectives that they can continue to build upon as they advance in their careers. Like many aspects of the PDP, we'll take time to do it carefully, thoughtfully, and collaboratively, not because it's how we always expect it to be done, but because we want you to have a foundational experience to draw on for many years to come.

Assessment is a complex, research-driven field in education. ISEE has narrowed the field down to three Focus Areas that are aligned with ISEE's other two themes:

1. Assessment as driver of iterative design and teaching

Within ISEE, assessment is considered part of designing an inquiry activity, and drives the design of an activity from beginning to end. PDP participants are expected to identify concepts that are core to the discipline they are teaching, then identify what kinds of difficulties learners have in using the concept in real world contexts -- for example, to explain a phenomenon they don't understand. PDP participants identify what it looks or sounds like when a learner understands versus when a learner has a misunderstanding or only a partial understanding. In addition, PDP participants identify an important cognitive STEM practice their students should learn. They identify what proficiency looks like in a way that is transferrable to a different context, and what kinds of difficulties learners may have with the practice. With learning goals clearly articulated, PDP participants use them to intentionally design their inquiry activity, working iteratively and returning back to review and revise their design. Then, when teaching, PDP participants maintain focus on the learning goals by monitoring learners' progress and making informed choices on how to help all students succeed.

2. Making learners' thinking visible

Facilitation sessions in the PDP emphasize the importance of using strategies to make learners' thinking visible to both the educator and the learner, as an important part of formative (ongoing) assessment. When learners are encouraged to explain or draw or show what they are thinking, they can more clearly pinpoint areas where they may need to improve their understanding, and facilitators can adjust their teaching to better support learners. As they design their inquiry activities, PDP participants develop a lesson plan that incorporates relevant strategies, and then they practice these strategies as they facilitate their activities. They make design choices that foster meaningful talk amongst peers, such as structuring investigation teams of 2-3 students, and actively monitoring social dynamics to support productive interactions within the teams. They learn on-the-fly facilitation moves for encouraging students to communicate their thinking in a range of ways, such as drawing, demonstrating with an experimental setup, or explaining. Once learners' thinking is made visible, the educator and learners can both monitor progress.

> **Box 2: Instructional move to make learner's thinking visible:** In an optics activity, the instructor asks learner to draw what she thinks is going on. This helps instructor see how the learner is thinking about light. Also, by creating a visual model, the learner may advance her own thinking or progress, or discover discrepancies on her own.

3. Assessing content understanding through learners' explanations

A powerful indicator of conceptual understanding is the ability to use a concept to explain a phenomenon, predict an outcome, or make an engineering design choice. PDP participants design an activity that has a culminating task requiring learners to explain their findings from their science investigation (or the solution to their engineering problem), by linking evidence they gathered in their investigation or design process to the core concept that is the intended learning goal. For example, PDP participants might have their students present a short, hand-drawn poster of their findings, introducing this task with a very carefully worded prompt that is intended to elicit the evidence needed to assess the students' understandings from posters and accompanying explanations. Participants use a scoring rubric to assess their students, prepared prior to teaching, and use it not just in the "culminating assessment task", but also to guide them as they formatively assess students' learning and give students feedback throughout the activity.

II. Assessment process and steps within the PDP

Assessment is integrated throughout the PDP and overlaps with many other aspects of the program, making it hard to disentangle. However, these are the phases or steps that are related to assessment within the PDP:

1. Determine a content goal/need (DTL generates a "Content Proposal")
2. Differentiate understanding from not understanding
3. Articulate a content prompt to elicit evidence
4. Determine a STEM practice goal/need
5. Design backward, including designing a culminating assessment task
6. Establish criteria for evaluating evidence (rubrics)
7. Gather evidence during activity
8. Evaluate evidence

1. Defining a content need and goal

As described in the Inquiry Theme document, all STEM fields have core, or foundational, concepts – concepts that have broad explanatory power (can explain many phenomena) and are tied to "big ideas". PDP participants will assess an aspect of a core concept that is identified by the Design Team Leader (DTL), who researches a specific need that is driven by the venue the activity will be designed for, and the identified difficulties learners have in coming to an understanding of the core concept. The DTL scans the literature, taps into their own teaching and learning experience, and learns about any existing assessment results to come up with a Content Proposal. The Content Proposal includes the proposed concept that will be the goal of the activity, why it is important, and what the DTL has learned about the particular needs and challenges in teaching the proposed concept (see Box 2 for an example). PDP instructors review the Content Proposal and help DTL refine it, and then the DTL discusses it with their team who will add any additional information from their own experience and/or knowledge of the literature around teaching and learning the proposed concept.

Box 2: Example of a PDP Content Proposal

Proposed concept goal: Intermolecular forces

Importance of concept: The concept of intermolecular forces is foundational and explains many phenomena not only in chemistry, but also in biology and other disciplines. For example, explaining phenomena like boiling points, solubility, and the structure of large bio-molecules.

Need that supports the choice of concept: A study of undergraduate and faculty, Loertscher et. al.[1] identified 5 core concepts in biochemistry and the particular difficulties that students have in understanding them. Intra- and Intermolecular forces was one of the five core concepts, and specific difficulties were identified

Difficulties in understanding: When students have not yet grasped the concept of intermolecular forces they:

- See interactions between molecules more about proximity than electrostatics, which comes out in representations of molecules interacting as:
    - An attractive interaction between neutral atoms
    - An attractive interaction between atoms with the same charge (or partial charge)

Loertscher, Jennifer, et. al. (2014) "Identification of Threshold Concepts for Biochemistry" CBE Life Sciences Education Vol 13. 526-528.

2. Differentiating Understanding from Not Understanding

When educators are assessing learners' understanding of STEM concepts, it is the educators' best judgment on whether the learner understands the concept, or does not yet have a sufficient understanding, based on evidence that comes from what a learner shows, says, writes, etc. For this reason, it is very useful for an educator to identify, in a very specific way, what understanding looks like versus what it looks like when a learner has a misunderstanding, a misconception, or incomplete understanding. PDP teams build from the work of the DTL's content proposal (Box 2), to make this distinction in the "Focusing on Content" session, and outline it in a way that makes sure the team has consensus and enables PDP instructors to give feedback. Box 3 shows an example, based on the concept of intermolecular forces.

Box 3: Evidence of understanding vs. not understanding for the concept of intermolecular forces.

When students have an incomplete understanding, or don't understand, they:	When students understand, they:
Show an attractive interaction between uncharged atoms *Show an attractive interaction between atoms with the same charge*	*Show an attraction between oppositely charged (or partially charged) atoms* *Show how the attraction affects orientation of individual molecules, and/or the structure of a large molecule*

3. Learning Outcome and Content Prompt

Armed with a good way to distinguish between learners understanding and not understanding, PDP participants are prepared to articulate a "content learning outcome" and a "content prompt." Teams are pushed to focus on <u>just one content learning outcome</u> for assessing their activity. Even though students will probably be learning many things, we have found from experience that it is most productive to focus on one (challenging) concept. It may also be that the learners in a PDP activity learn and use a number of concepts along the way, but it is still most effective for PDP participants to identify and assess the concept that best represents an enduring understanding. This helps to keep the activity design tightly focused, and makes the final judgment of whether learners "got it" a feasible task for PDP participants.

PDP participants define a "content prompt" that is intentionally designed to elicit evidence needed to make the distinction of whether or not learners "got it". The content prompt does not leave to chance that learners will show or tell the instructors what they are looking for. It is carefully crafted and fine-tuned to make sure learners provide the instructors with the evidence of understanding that needs to be gathered. The format of the assessment task (e.g., poster, jigsaw, etc.) can be determined at later stages of the design, but the actual wording of the content prompt is very important before beginning to design an activity. Note in Box 4 that the content prompt includes "at a molecular level." This part of the prompt is a part of eliciting the key evidence instructors are looking for. In this example the evidence includes drawings of molecules, and the instructors worded the prompt to be explicit about that.

> **Box 4. Examples of learning outcome and content prompt for the concept of intermolecular forces**
>
> **Content learning outcome:** *Students can explain phenomena observed when various materials are interacting, using intermolecular forces.*
>
> **Content prompt (e.g. as part of directions given to students on what to present on a poster):** Explain the phenomenon you observed at a molecular level

The content prompt will serve as the guidepost for PDP participants to design toward. Later, as teams move toward teaching their activity, it will help participants plan for their on-the-fly teaching moves, and engage in ongoing, informal assessment of learners' progress. Finally, the content prompt will become part of the assessment task and will give PDP participants evidence of learning, which they can then use to make inferences about learners' progress, and have a concrete way to evaluate their design and teaching.

4. Culminating assessment task

Assessment tasks are assignments (small or large) given to learners that are designed to provide evidence that will allow an educator to assess learners' knowledge or skills. More traditional assessment tasks are multiple-choice tests, fill-in-the-blank questions, or problems in which the learner shows their work. These kinds of tasks are often separated from teaching and learning, and are usually more contrived. However, one can also use an "authentic assessment task" that is part of the learning process and applies knowledge and skills to a real-world challenge. In authentic assessments, students are still learning at the same time as they are being assessed. In the PDP, we focus on authentic assessment tasks, for example having learners explain what they learned from an investigation in a poster constructed at the end of their investigation time. An authentic "culminating assessment task" is considered an essential component of a PDP inquiry.

PDP teams all include an authentic assessment task in their activity that is designed to create an opportunity for the instructors to assess their learners and simultaneously continue the learning process. Common assessment tasks are poster presentations or jigsaw discussions at the end of the investigations, but these are not the only formats that could be used. The assessment task is assigned to students through a carefully crafted set of prompts that elicit the key evidence needed to assess learners, and ultimately for PDP participants to evaluate their activity design and facilitation.

There are two important aspects of a culminating assessment task that need to be considered when designing this component: 1) the specific wording of key prompts, such as the "content prompt" already discussed above; and 2) format of overall culminating assessment task. At the Inquiry Institute, PDP participants will brainstorm different kinds of culminating assessment tasks, and the tradeoffs involved in them, so that as they begin designing, they have a few options to choose from. The format of a culminating assessment task should be carefully designed, using ISEE themes as well as considerations from applying the How People Learn framework. Planning how the culminating assessment task is introduced and wrapped-up to learners is extremely important, along with timing, and how it is facilitated. PDP participants should carefully consider the backgrounds of their learners and how they will engage with this part of the activity. For example, how will instructors make sure that all learners get recognized for their contributions? Or, if learners will be asked challenging questions, is it likely that some will have experienced this and be comfortable, and others will not know this STEM norm and take it much more personally than is intended?

5. Defining a STEM practice goal/need: In the "Improving STEM Practices" session at the Inquiry Institute, participants will gain experience articulating STEM practices as the kinds of authentic reasoning processes used by scientists and engineers. During this session participants consider core STEM practices (e.g. designing investigations) and more specifically defined aspects of a practice, which illuminates the evidence that should be looked for in evaluating learners as they engage in a particular practice. Participants will be pushed to define <u>one main STEM practice</u> that they will focus on, even though their learners will likely engage in a wide range of practices. This will help participants focus deeply on one practice, and the subtle, nuanced ways that scientists and engineers use the practice within their field of study. In the Understanding by Design reading, the "conceptual or strategic element of any skill (*practice)*" brought up as a curricular focus that meets the criteria for an enduring understanding is a more generalized way (not specific to science or engineering) of describing the STEM practices focused on in the PDP.

Box 5. Sample STEM practice
Designing experiments. Some examples of more specific aspects of this practice:
- *Identifying variables in complex systems*
- *Devising controls*
- *Devising ways to make measurements*

Box 5 gives an example of a couple of specific aspects of the core practice of designing investigations. For example, variables are not always obvious so identifying them is often quite challenging. Similarly, there is often not a straightforward way to make a desired measurement, and it takes reasoning and creativity to devise an indirect method of measuring something. It should be noted that there are many other specific aspects to designing investigations – the three in this example represent a set appropriate for focusing an activity on. PDP participants identify a core STEM practice, and a few specific aspects of it that will help them design and teach their activity, and make focused observations of learners using the practice.

6. Creating a rubric

Teams define the criteria that they will use to differentiate "understanding" from "not yet understanding" the content learning outcome and create a "rubric", or a tool for scoring learners' level of understanding (see Appendix). The final rubric draws from early work on the "content proposal" as well as follow-up work by the design team, and may be very basic with a binary choice between understanding and not understanding, or could have more levels such as a 3-point scale. Because PDP teams are piloting a new activity (including a new culminating assessment task) an elaborate rubric isn't usually feasible. In addition, PDP participants are encouraged to keep it simple and just focus on a couple of key aspects of a concept. The criteria outlined in the rubric can be in a narrative form, or may reference a diagram. Before teaching, team members may want to add some additional notes that relate to the specific station or investigation that they are facilitating. It is very important that each team member knows what she/he is looking for when assessing their learners' knowledge.

An important aspect of using a rubric is accounting for the fact that an educator may not get the evidence needed to make an inference about whether learners understand the intended learning outcome. For example, the content prompt may be misinterpreted by learners, or it may be too vague to elicit the key evidence. Alternatively, the culminating assessment task may have more prompts than learners can practically respond too, so they are forced to pick and choose, and skip the content prompt. All of these situations reflect a weakness in the assessment, which quite different than a learner leaving out something important due to an incomplete understanding of a concept. For this reason, PDP participants are asked to include a "lack of evidence" column for all their rubric dimensions, so that after teaching, they can reflect on the evidence they gathered, or if they didn't gather it, why they may not have gotten what they need to assess their learners.

Some important notes:

- Note that the rubrics do not simply say something like "molecules drawn correctly," or "molecules shown interacting with correct orientation." The rubric specifies what "correctly" is, and specifically what the instructor is looking for.
- The rubric should apply to all of the investigation areas learners engaged in, and could likely be used in other contexts in which the learning goal was intermolecular forces

> **Common pitfall:** ***"Our team has a different rubric for each of our 3 investigation areas."*** If your team is creating a different rubric for each investigation area, it most likely indicates one of two things:
>
> - You might have three different inquiries (each investigation area has different content goals)
> - You haven't (yet) found the common, or generalizeable "evidence of understanding." This often takes stepping up a level from the specifics of a station to identify what is common across the investigation areas. Note in the above chemistry example specific chemical compounds (e.g. benzoic acid) are not used.

7. Gathering evidence during an activity

Throughout the activity facilitators are eliciting and gathering evidence of their learners' understanding, and using it formatively to guide teaching and learning. During the culminating assessment task (e.g. posters), facilitators will gather evidence on the learners that they specifically facilitated, taking enough notes that an evaluative score can later be given. Teams can decide whether they will be considering teams or individuals. Each facilitator will also have a lot of evidence of learners' understanding in his/her head, and possibly in notes, having spent hours working with the learners. The evidence gathered during the assessment task may or may not match with this informally gathered evidence, and will make a productive point of discussion in the debrief. As PDP participants teach their activities, they may find that some unanticipated evidence of learning arises. This too is important to note and debrief about, as it could be an important source of assessment if the activity is taught again in the future.

Participants will also observe how learners engage in the STEM practice, using a second rubric prepared on their STEM practice (see Appendix). PDP participants are not asked to create a culminating assessment task for the STEM practice, but are asked to at least make observations and take notes that can later be used to reflect on how learners demonstrated proficiency, and what aspects of the practice were most challenging.

8. Evaluating evidence

Each facilitator will use their content rubric and score the learners that he/she specifically facilitated, based on the evidence gathered in the assessment task, noting any evidence that was different than what they had anticipated. Facilitators will bring a record of their scores to the team's post-teaching debrief, so that the whole team can discuss how well the assessment task matched their informally gathered evidence, and use the results of both to evaluate their design in relation to the intended learning goals, and/or in relation to unanticipated learning outcomes. Teams can also brainstorm ideas for improving the assessment task or doing something completely different.

9. Reflecting and Reporting

Finally, teams will evaluate their design based on their assessment. Teams may find that the learners actually learned something different than intended, or the learning goals may have been too ambitious so very few learners got to where the PDP team expected. Or teams may find that their learners really hit the mark during investigation time, but when they presented their posters left off the most important things. These kinds of situations are to be expected in the first implementation of an inquiry activity, and can be disappointing for PDP participants. However, the most important part of the PDP experience is reflecting and learning from the design and teaching experience. PDP participants who can see both the strengths and flaws in their designs, and can make informed suggestions on how their design can be improved to accomplish the PDP task are considered to be successful.

The final product of PDP participants' work is a lesson plan with revisions that integrate what was learned from teaching, assessing, and then evaluating; as well as a post-teaching report that includes "assessment summaries" of students' learning with regard to the main content and practice that the inquiry activity focused on (see Appendix for examples). Reflecting and identifying how a design could be improved to accomplish goals, or identifying a relevant but unanticipated learning outcome, are successful outcomes for PDP participants. At the end of the PDP experience, we want you to be able to articulate learning goals, determine evidence of learning, gather evidence, and use that to critically evaluate a design. It is the reflection and what you learned along the way that we value most.

APPENDIX: PDP Assessment Reporting

Below is an example of the information participants are expected to document in their lesson plan with respect to the main content goal of their activity and their content rubric.

Core concept: intermolecular forces

Content learning outcome: Learners will use the concept of intermolecular forces to design a drug molecule to optimally bind in a target drug receptor site.

Content prompt: "Show at a molecular level why the molecule you've chosen is 'best'"

Culminating Assessment Task: Jigsaw format at the end of the activity.

Content Rubric:

Dimensions: **Components or "knowledge statements"**	**M** (evidence needed to make a judgment is missing)	**0** (evidence that learner has misunderstanding or incomplete understanding) **Diagram with representations of molecules shows…**	**1** (evidence that learner has sufficient understanding) **Diagram with representations of molecules shows…**	Score
Molecules can have full, partial and/or momentary charges		Show a charge on non-polar region Show a + charge where it should be – or vice versa	Polar regions, and partial, full, or momentary charge	
Attractive interactions between molecules are based on opposite charges attracting at specific regions		Molecules oriented so that regions with same charge are interacting	Molecules oriented so that regions with opposite charges attract	
There are different types of interactions between molecules, with different strengths, often in competition		Identifies only one interaction, though other are present	Identifies regions with different interactions and relative strengths (H-bond > dipole-dipole > dispersion)	
TOTAL SCORE				

PDP participants are also expected to summarize their assessment findings in their post-teaching report.

Example content assessment summary:

Learners were assessed based on how they used the concept of intermolecular forces to defend their choice of a drug molecule. A rubric was created, and learners were scored based on whether they were able to 1) identify polar regions of a molecule; 2) show molecules oriented so that opposite charges attract; and 3) and identify multiple type of interactions between two molecules and their relative strengths. Learners demonstrated their understanding as they justified their choice of a drug molecule in a jigsaw format, and supported it with evidence from their iterative design process. 12 of the 15 learners scored 2 out of 3 total points, which was what we considered proficiency. Only two of the learners scored a point for identifying the different types intermolecular interactions and their relative strengths, but this was due to a lack of opportunity to demonstrate this understanding. Future versions of this activity would need to make sure that there were multiple types of competing interactions, and that learners are prompted to include this as they report their findings. In addition, instructors noted that many learners were unclear on the difference between covalent and non-covalent interactions, and future versions of this activity could adapt the design and rubric to anticipate this learner difficulty.

As with content, participants are expected to document the STEM practice their activity focuses on and a practice rubric in their lesson plan. An example follows.

Core practice: Designing experiments

Practice Rubric: From a "rubric for experimental design" published by Dasgupta, et. al. (2014) CBE-Life Sciences Education, Vol 13, 265-284

Aspects:	Lack of evidence (did not observe learners enough to decide between A and B)	A. Evidence of difficulty: what it looks like when a learner needs to work more on the practice	B. Evidence of proficiency: what it looks like when a learner is proficient with the practice
Motivated by a hypothesis		Only refers to a dependent or independent variable. Does not clearly indicate the expected outcome to be measured from a proposed experiment	Hypothesis is a testable statement, with a predicted relationship between a dependent and independent variable
Experimental sample or treatment group		Haphazard assignment of treatments to experimental samples, inappropriate to goal of experiment Experimental conditions do not yield results aligned with the goal of the experiment	Experimental samples exposed to experimental conditions that vary in a specific way
Controlling outside variables		Variables unrelated to research question are mismatched across treatment and control groups	Control and treatment group matched as closely as possible, to reduce effect of unknown variables

PDP participants summarize their assessment findings with respect to the practice in their post-teaching report, too.

Example practice assessment summary:
To assess learners' proficiency with designing experiments, learners were observed during their team investigation time as they carried out their own experiments to explore their questions related to intermolecular forces. Instructors used a rubric as a formative assessment tool, which identified evidence of proficiency and difficulty related to three specific aspects of experimental design: 1) articulating a testable hypothesis; 2) having a treatment group; and 3) controlling outside variables. Instructors observed that many students had difficulty with the three aspects of experimental design noted above, but with support from the instructors most were able to demonstrate proficiency with the first two aspects. The third aspect, controlling outside variables, was difficult to assess because the software used in the activity already eliminated outside variables, so learners were not faced with this challenge. Instructors noted that learners had more difficulty than anticipated in articulating a testable hypothesis. For example, some learners did not have a predicted association between the treatment and outcome variable -- some only referred to the treatment or the outcome variable, but not both. Future instructors of this activity should consider using these aspects in the assessment, and may want to design an activity component to help students who are struggling. For example, early on in the investigation time teams could be paired for a simulated group meeting in which each team shares their hypothesis, and the other team asks questions and recommends improvements in the way that the hypothesis is worded.

Acknowledgments:
This material is based upon work supported by that National Science Foundation (NSF) under Grants: AST#1347767 and DUE#1226140.