INSTITUTE for SCIENTIST & ENGINEER EDUCATORS

Assessment-Driven Design: Supporting Design, Teaching, and Learning

Lisa Hunter, Rafael Palomino, Barry Kluger-Bell, Scott Seagroves, and Anne Metevier

Suggested citation: Hunter, L., Palomino, R., Kluger-Bell, B., Seagroves, S., & Metevier, A. (2022). Assessment-Driven Design: Supporting Design, Teaching, and Learning. *UC Santa Cruz: ISEE Professional Development Resources for Teaching STEM (PDP)*. Retrieved from https://escholarship.org/uc/item/2n40d3kz

This paper was written and produced by the developers of the Professional Development Program (PDP) at the Institute for Scientist & Engineer Educators (ISEE) at University of California, Santa Cruz. The PDP was a flexible, multi-year program which trained participants to teach STEM effectively and inclusively at the post-secondary level. Participants were primarily graduate students and postdocs pursuing a broad range of science and engineering careers. Participants received training through two in-person multi-day workshops, worked on a team to collaboratively design an authentic, inclusive STEM learning experience (an "inquiry" lab), and then put their new teaching skills into practice in programs or courses, mostly at the college level. Throughout their experience, PDP participants used an array of online tools and received coaching and feedback from PDP instructors. The overall PDP experience was approximately 90 hours and was framed around three major themes: inquiry, assessment, and equity & inclusion. Leadership emerged as a fourth theme to support PDP teams, which were each led by a participant returning to the PDP for a second or third time, who gained training and a practical experience in team leadership. ISEE ran the PDP from 2001-2020, and there are more than 600 alumni.

CONTEXT FOR THIS PAPER WITHIN THE PDP

This paper outlines one of the PDP's three major themes, "assessment-driven design," that participants use throughout their experience. The paper was read by participants prior to beginning their PDP experience, and then served as a reference throughout the program. It is primarily focused on using assessment to design an authentic and inclusive STEM learning experience, and in particular learning STEM content. Other PDP resources focus on assessing STEM practices.

The PDP was a national program led by the UC Santa Cruz Institute for Scientist & Engineer Educators. The PDP was originally developed by the Center for Adaptive Optics with funding from the National Science Foundation (NSF) (PI: J. Nelson: AST#9876783), and was further developed with funding from the NSF (PI: L. Hunter: AST#0836053, DUE#0816754, DUE#1226140, AST#1347767, AST#1643390, AST#1743117) and University of California, Santa Cruz through funding to ISEE.

Assessment-Driven Design: Supporting Design, Teaching, and Learning

Lisa Hunter, Rafael Palomino, Barry Kluger-Bell, Scott Seagroves, and Anne Metevier (2022)

We acknowledge the work of many past PDP collaborators.

I. Overview

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. In the PDP, participants get into the details and nuances of the backward design process, which we call "assessment-driven design" (though either is fine – it is the process that is important). We use assessment-driven design throughout the PDP, including the integration of a culminating assessment task into all PDP-designed activities.

PDP participants generate a set of three tools that guide the design process, are refined throughout the design process, and ultimately become the tools used to assess their learners' success in using a concept to explain, design, or predict something, both throughout and after their teaching. Those tools are:

- **Assessment Prompt:** A carefully crafted, concise directive to the learners that conveys how they will demonstrate their understanding of a concept to explain, design, predict, or justify something.
- **Learner artifact (or learner work product):** Something created by individual learners, in response to the assessment prompt, that allows you to assess their performance towards the intended learning outcomes defined in the rubric. The artifact includes drawings, text, or other ways of demonstrating they have achieved the learning outcome.
- **Rubric:** A common assessment tool that helps you articulate your expectations for learners and define what will be assessed, and the criteria for how it will be evaluated.

The three tools above (described in more detail below) drive the design process and are a prominent part of the PDP experience in the early phases. By the end of the PDP, we want participants to evaluate the degree of success that their design and teaching had in achieving the goals for their learners, using evidence gathered formally in an authentic assessment task and informally throughout their teaching. It is the reflective process that we value most and expect that by the end of the PDP experience participants will have practiced the backward design process in a concrete way, and used it to evaluate their design and teaching. We hope 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.

Disentangling STEM content and practices: Teaching, learning, and assessing STEM practices (e.g. designing experiments) is an important part of the PDP. This document is focused primarily on assessing content, but because scientists and engineers learn and apply conceptual understanding through practices, it is important to consider those practices in relation to content. 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).

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 they conduct 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.

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 an authentic assessment 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. This is quite different, and much more challenging, than having students state or define a concept. That is, many students are able to memorize and repeat a concept, but are unable to use the concept in a real-world context.

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 assess an aspect of a core concept that is identified by the Design Team Leader (DTL). The DTL researches a specific need that is driven by the venue the activity will be designed for, and then identifies difficulties learners have in coming to an understanding of that core concept. At the Inquiry Institute the DTL shares this work which is documented in a "Content proposal" (see Box 2 for an example). DTLs are responsible for decisions related to the core concept, in consultation with the venue lead and PDP instructors.

Teams are pushed to focus on just one content learning outcome 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 is important for learners to develop an enduring understanding of. This helps to keep the design tightly focused and makes the final judgment of whether learners "got it" a feasible task for PDP participants.

Box 2: Excerpts from an example PDP Content Proposal

Core concept: 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.

II. Tools for Assessment-Driven Design

There are three primary assessment tools used in the design process to drive decision-making throughout the process of developing the inquiry. The **rubric**, learner **artifact**, and **assessment prompt** are described in more detail below. Examples are based on the core concept of intermolecular forces (see Box 2 above).

1. Rubric

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 specific way, what understanding looks like versus what it looks like when a learner has a misunderstanding, a misconception, or incomplete understanding. Included within the DTL's content proposal (Box 2) is something of a first-draft of a **rubric**, which is further refined by the whole team before and during the Inquiry Institute to create a draft that is used in the design process. Box 3 shows an example of a rubric that a team might create (based on DTLs first draft) and that could be used in the design process. It should be noted that the rubric is refined to enable scoring later, when the team is further along in the design process.

Box 3: Example of a rubric sufficiently detailed to use in the design process

Dimensions of concept	When students have an incomplete understanding, or don't understand, they say or show: Diagram with representation of molecules shows…	When students understand, they say or show: Diagram with representations of molecules shows…
Molecules can have full, partial and/or momentary charges	A charge on non-polar region 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 others are present	Identifies regions with different interactions and relative strengths (H-bond > dipole-dipole > dispersion)

2. Assessment Prompt

PDP participants create an **assessment prompt** that is intentionally designed to elicit evidence needed to make the distinction between learner understanding and not understanding. The assessment 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 authentic assessment task (e.g., poster, jigsaw, etc.) can be determined at later stages of the design, but the actual wording of the assessment prompt is important to consider before beginning to design an activity. Note in Box 4 that the assessment prompt includes "at a molecular level." This part of the prompt is critical to elicit 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. Also note how similar the assessment prompt and content learning outcome are; the difference is just in whether it is phrased as a directive to the learner, or in describing the intended outcome to others.

PDP participants articulate a content learning outcome and an assessment prompt. Examples of these statements are shown in Box 4. "Learning outcome" is a term used throughout educational systems and contexts. While "assessment prompt" has specific meaning in PDP, its purpose will most likely be understood in a general way by educators in other

Box 4. Examples of learning outcome and assessment prompt for the concept of intermolecular forces

Assessment prompt (e.g. as part of directions given to students on what to present on a poster): Explain the findings of your experiment at a molecular level, describing the interactions between molecules.

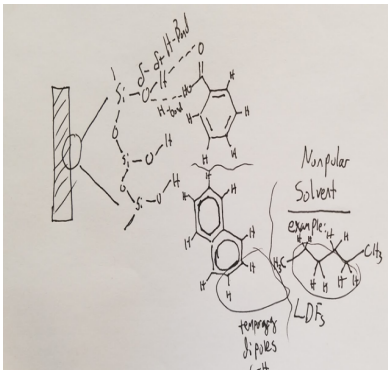
Content learning outcome: *Students will explain a chemical phenomenon at a molecular level using the concept of intermolecular forces.*

contexts. For the design process in the PDP, we have found the assessment prompt to be a more useful to focus on than the learning outcome.

3. Sample Learner artifact

A learner **artifact** (or learner work product) is something created by individual learners, in response to the assessment prompt, which allows instructors to assess learner performance towards the intended learning outcome. The artifact includes drawings, text, or other ways of demonstrating learners have achieved the learning outcome. The artifact should be scored later, like any other assignment a student turns in. Additionally, an artifact can be used in combination with scoring learners' oral presentations (though this is very hard to do on the fly, and especially with a newly designed activity, and not recommended for PDP participants). Importantly, PDP participants should assess the artifact using their **rubric**. Ideally, the artifact should include all the evidence needed to assess every dimension of that rubric. DTLs are asked in the content proposal (Box 2) to show example(s) of learner work. Building on these initial ideas, at the Inquiry Institute teams will generate sample artifacts that their learners might produce. By producing sample artifacts - including one that represents understanding and one that represents misunderstanding - teams come to consensus and establish the design requirements of their activity. In other words, if the desired learner artifacts are the end goal, what must the assessment prompt, rubric, assessment task, and design do to give learners the opportunity to demonstrate their understanding in an artifact.

Box 5: Example sample learner artifact



The three assessment tools -- rubric, assessment prompt, and artifact -- serve as the guidepost for PDP participants' design efforts. Teams go through multiple iterations of design using these tools to make design choices and refinements. Over time the tools get refined as more is learned through iteration, and will be used to assess learners.

NOTE: *As part of the iterative backward design process, participants brainstorm possible investigation pathways and assessment tasks as a means of developing the tools. The early brainstorming (Day3 and 4 of Inquiry Institute) is just enough to get some context for developing the three tools. It is very difficult to create the three tools without some context. In these early brainstorming sessions, teams should just get something on the table to work with and should not be concerned that it's not quite right, and should not get attached to what is brainstormed. There is much more time for designing at the Design Institute.*

III. Assessment-Driven Design

The three assessment tools described above enable PDP participants to move forward efficiently and effectively in the assessment-driven design process. The first components designed are the **culminating assessment task** and the **investigations,** which are each

described below. After some initial design on these two components, it is a good time to consider STEM practices, also described briefly below.

1. 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 learning at the same time as they are being assessed, and they demonstrate their understanding in a way that mirrors the way it is done in the real world (e.g. by a practicing scientist or engineer). In the PDP, we focus on authentic assessment tasks, such as 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.

Box 6: Culminating assessment task for intermolecular forces example activity.
In the investigation phase, students were tasked with trying to understand how different molecules interact with each other in teams of three. In the culminating assessment task, those students could be split up from their investigation teams, and asked to present their team's results in new groups with other students who investigated different phenomena (a "jigsaw"). To prepare students for this task, they could be asked to respond to a prompt, which includes the assessment prompt, in addition to other instructions for presenting their results. This individual student artifact is used to help the students present their results, but it is also collected by the instructor to be assessed with a rubric. This highlights the importance of a tight alignment between the rubric, the assessment prompt, and the type of student artifact asked for. All of these pieces are part of the culminating assessment task.

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.

From a design standpoint, there are three important aspects of a culminating assessment task that need to be considered when designing this component: 1) the specific wording of the **assessment prompt** already discussed above; 2) format of and contexting to learners about the overall culminating assessment task; 3) how to collect something that can be scored after the activity (an **artifact**). An example is given in Box 5. The format of a culminating assessment task should be carefully designed, using ISEE themes as well as considerations from applying the *How Learning Works* 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 other will not know this STEM norm and take it much more personally than is intended? Finally, PDP participants should figure out a good way to gather an artifact, such as something written, from individual learners. This often provides a way for learners to reflect, and continue learning, which is an added benefit.

Investigations
The investigation phase of the inquiry activity is usually the main part of the activity, and learners spend the most time in this component. Investigation time provides an opportunity for learners to plan and carry out their own way to produce evidence that will support their findings. There are many resources describing investigations, and how to design them (e.g. Design Notebooks), which are not included here. In relation to assessment-driven design, it is important to consider the design requirements of the investigation components. In other words, what must the investigations do in order to get learners to respond to the assessment prompt and produce the desired artifact. There should be strong continuity between the Investigation and CAT, in that learners have been gathering evidence for the CAT all along during the Investigations.

Integrating a STEM practice into design STEM practices are covered in other documents and will not be covered in detail here. However, the integration of STEM practices is a key part of the design process and is discussed here in relation to how consideration of STEM practices folds into the design process. From years of experience, the PDP has found that the most efficient design process is to start with content, do an early iteration of design, and then bring in goals related to STEM practices. 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. At this point in the design process PDP teams can integrate STEM practices into their design considerations.

IV. Using Rubrics for Assessment

A rubric is common assessment tool that helps articulate expectations for learners and define what will be assessed, and the criteria for how it will be evaluated. In the PDP, the DTL creates the first draft of the rubric, which drives design. The final rubric draws from this early work and the refinements made by the team during design iterations. This rubric may be basic, with a binary choice between understanding and not understanding, or it could have more levels, such as a 3-point scale. Because PDP teams are often piloting a new activity (including a new culminating assessment task) an elaborate rubric isn't usually feasible. 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 they are 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 to, forcing them to pick and choose, and potentially skip the content prompt. All of these situations reflect a weakness in the assessment, which is 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 space for "lack of evidence" in 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 criteria outlined in 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:

1. You might have three different inquiries (each investigation area has different content goals)
2. You haven't (yet) found the common, or generalizable "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.

1. Gathering evidence for content understanding through artifacts

Throughout the activity, facilitators are eliciting and gathering evidence of their learners' understanding, and using it formatively to guide teaching and learning. The culminating assessment task (e.g. posters, jigsaw, etc.) should include the collection of an artifact that can be later scored using the rubric. Artifacts can take on many forms, such as a written response to a prompt(s), a draft of a poster, or even an actual poster, but should enable the

facilitators to assess individual learners' understanding. Each facilitator will also have evidence of learners' understanding in their 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. Importantly, while the informal evidence that is gathered is important and of value both for instructors and the learners (assisting in their own understanding of the content), the creation of an artifact leads to a permanent record of what a learner understood that an instructor can assess, even if they've missed informal evidence of understanding. 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.

2. Assessing STEM Practices

Participants will use their STEM practice rubric to formatively assess their learners during the activity. PDP participants are not asked to create a culminating assessment task for the STEM practice, but are asked to 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. However, participants may want to devise a summative assessment of the practice, especially those that have done the PDP at least once already and who want to gain a new experience. PDP participants interested in designing a summative assessment of the STEM practice should consult with PDP instructors to get some ideas on how to go about this.

3. Evaluating evidence

After the activity, PDP participants should score the artifacts. Ideally, time is scheduled immediately after the activity to score the artifacts. Each participant scores all the learners' artifacts. Teams may want to use an average score for reporting back to ISEE. During the debrief, participants 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.

4. 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.

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 of what you learned along the way that we value the most.

Appendix A: Some Introductory Ideas and Terminology Related to Assessment

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.

Two major types of assessment

- **Summative assessment:** assessment *of* learning; "final" judgment; generally no chance for revision or affecting student learning. When a student takes a final exam, this is a summative assessment.
- **Formative assessment:** assessment *for* learning; ongoing judgments; generally with chances for revisions and new effects on student learning. When a student turns in an assignment, gets feedback from the instructor, and then can make revisions, this is formative assessment.

Purposes of assessment may include: Research and/or evaluation of programs, curricula, schools; labeling/sorting students; diagnosis of students; evaluating teachers; supporting learning.

In the PDP, assessment is used:

- For design
- For learning: to help your learners
- To evaluate your design and teaching
- To give you a concrete experience in designing authentic assessment
- Collectively, your assessment helps us to build a case for the effectiveness of the PDP

Goals Assessment is in relation to specific goals. Goals can be collective goals for a cohort of learners (e.g. creating a learning community, increasing persistence rates), or individual. Individual goals, or intended "learning outcomes" are commonly broken into three categories:

- Content (what students know)
- Process (what student can do)
- Attitudinal (or "affective", for example, developing an identity as a science person)

Validity (sometimes the similar term **authenticity**): A valid assessment really measures what it sets out to measure. For instance: Does a timed essay test measure understanding or just how fast the students can write? This is a question of validity.

Reliability: A reliable assessment yields consistent results when applied repeatedly to similar students.

You can think of these terms as being very similar to "accuracy" and "precision" in experimental measurement. Just as a measurement can be *precise* without being *accurate* and vice-versa, so an assessment can be *reliable* without being *valid* and vice-versa.

Rubric
Guidelines which state the characteristics/dimensions being assessed and give performance criteria with a rating scale – a scoring *guide*. *Not* an absolute, rigid mechanism but a *guide* for making judgments.

Features or parts of a rubric:

- **Dimensions / elements / characteristics / evaluative criteria:** the breakdown of dimensions on which performance is judged.
- **Quality definitions / performance definitions:** descriptions of performance at different levels.

Assessment Prompt
Broadly, an assessment prompt is a carefully crafted, concise directive to the learners that provides an opportunity to check student learning. The PDP assessment prompt is written in a way that asks learners to use evidence from their investigation to demonstrate their understanding of a concept to explain, design, predict, or justify something.

- Minor variations in the wording of the assessment prompt turn it into
    - The goal of the activity given to learners at the beginning of the activity
    - The prompt given to learners at the beginning of the investigation component
    - The content learning outcome

A good assessment prompt directs learners to put information in the artifact generated during the CAT that can be assessed by the rubric towards finding out if learners were able to apply the stated learning outcome, and all dimensions of the rubric.

Learner Artifact (Learner Work Product)
Created by individual learners, in response to the assessment prompt. Allows you to assess their performance towards your intended learning outcomes.

Includes drawings, words, or other ways of demonstrating they have achieved the learning outcome.

Used on its own to assess learners' understanding, or in combination with scoring learners' presentations or other tasks. Importantly, the artifact should be able to be assessed by the **rubric** that you have created, and ideally should elicit all evidence needed to assess all dimensions of that rubric.

Appendix B: Frequently Asked Questions About PDP Assessment

How do we decide what goes in the dimensions? Dimensions should be 2-3 aspects of the content (or practice, in the case of the practice rubric), and should be a few things that are key to differentiating "understanding" from "not yet understanding." Dimensions are not steps in a process, but may be components of a chain of reasoning. Some content rubrics have dimensions based on claim, evidence, and reasoning (see example #5 in the appendix). They should be defined in a way that goes beyond the activity, to support learning a concept (or practice) in a way that is transferrable to new contexts.

Is it better to have just a binary scoring scale (0 or 1) or more levels? This is completely up to teams to decide. Some find it easier to have fewer choices; others find that they need to have a place for something in between (much like a "partial credit" option when grading). The important part is to make the rubric something your team can use, and which would provide a means of giving useful feedback to learners.

What is the difference between a "0" and "M" (missing evidence)? Lacking enough evidence to evaluate learners can arise from a number of situations. It is important for educators to be clear on when learners leave out something because they don't understand ("0"), versus when there is a problem with the activity, the rubric, or the method of gathering evidence ("M"). You might decide that you need to throw out a dimension with a lot of "M"s because you see (after scoring) that there was a design flaw or a problem with your culminating assessment task that did not elicit the evidence you needed.

How and when do we do the actual scoring? Teams should score learners' artifacts from the culminating assessment task. **PDP participants must have their learners create an artifact** that can be scored later (it is too difficult to try this on the fly when you are teaching an activity for the first time).

Who should score? All PDP participants should score each artifact.

Do we share the rubric with our students? This is left up to participants. Participants should definitely share their goals/learning outcomes with learners, but may want to go further and share the dimensions of the rubric, if that does not take anything away from the learning process. Participants could consider sharing a modified version of the rubric.

We have two phases to our activity, which do we score? Participants should assess the core concept that is learned from the activity. If the team anticipates that the learners will have learned as much as they can at the mid-point of an activity, careful thought should be given to what is going on in the second half. One possibility is that the team is/has inadvertently designed two activities with two sets of goals.

Appendix C: Characteristics of a useful rubric

- There should only be a few dimensions that are brief and focused on a single aspect of the content (dimensions do not need to represent everything about a concept -- just a few key things)
- Dimensions can be chains of reasoning for the concept or specific aspects of the concept that are not interdependent, chosen because they are aspects that illuminate the learners' understanding
- Dimensions should be generalizable beyond the inquiry activity. You could use the rubric in another activity with the same learning outcome but in the context of a different lab or assignment.
- Rubric quality definitions should differentiate understanding from not understanding, and gradations in between, if necessary. Quality definitions should describe the concrete evidence of what learners says, does, draws, or otherwise demonstrates various levels of understanding
- Rubrics should be designed to be valuable for students, as well as instructors. If given to a student with the assignment description, would it help the student monitor and assess their progress towards your learning outcome? How does the rubric express or frame understanding without replacing the learning process (i.e. by plainly stating the knowledge students should gain or use)? If you returned an assignment with a rubric, could the student recognize their strengths and weaknesses to help them direct their effort towards improving?

Some issues or points to watch for with rubrics:

- The dimensions / elements should be *teachable* – students should be able to learn and demonstrate this knowledge by engaging in your activity
- Watch out for *adjectives* and *adverbs* (like "correctly" in the example within the text) that should be "unpacked" to reveal what you are really looking for.
- Rubrics should be *specific* enough to be really useful – vagueness clarifies little. But, on the other hand, the rubric can never be so specific that it encompasses all that any students might ever say/do – the teacher must still make the judgments, using the rubric only as a *guide*.

Appendix D: Common challenges/pitfalls with Culminating Assessment Tasks:

- Culminating assessment task (CAT) prompts are very guided (e.g. like filling in blanks on a worksheet), and not very authentic - it is hard to do, but work carefully to craft prompts to get what you need to assess, but not lose authenticity
- Teams create two CATs: one that creates an artifact and one that has another way of reporting (e.g. posters). If multiple ways of explaining/reporting are solicited, they should be the closely related and based on the same assessment prompt.
- Assessment prompt is missing (or appears to be missing from the CAT prompts)
- CAT asks for a lot, so learners will likely have to choose what to show, and then can avoid addressing the most important prompt (the assessment prompt)