The Evidence of Learning and Impact Framework: A Delphi Study

Authors:

Kirstin Moreno, M.S.Ed., Ph.D.
Oregon Health & Science University

Sarah Jacobs, M.Ed.
Clark College

Constance Tucker, M.A., Ph.D.
Oregon Health & Science University

ABSTRACT

Learning outcomes frameworks are useful in course, program, and institutional assessment as well as continuing education or professional development contexts and help ensure that different aspects of learning are addressed. This article describes a Delphi study conducted to iterate and improve on the authors' novel Evidence of Learning and Impact Framework using assessment experts' feedback. This new framework is useful broadly within adult and higher education and uniquely incorporates an emphasis on attending to the impact of the learning on the learner, the impact of the learning on others, and encourages the use of equity lenses when examining learning. Niederberger and Spranger (2020) encourage more transparency about Delphi techniques used in scholarship, so the authors also provide many details about the Delphi process used. We hope that the Evidence Framework will challenge educators to think differently, more broadly, and more deeply about the kinds of learning they foster and assess.

Correspondence E-mail: kirstin@morenos.name
Keywords: Learning Outcomes Framework; Delphi Technique; Educational Assessment; Equity-Centered Assessment; Higher Education Assessment

Orienting to Our Previous Work

Since 2020, the authors have been engaged in a multi-step process to design a learning outcomes framework that better fits our institution's needs, and that also aligns with the assessment culture we want to promote in light of calls for incorporating antiracist and equity-focused approaches into teaching and learning (Alegría et al., 2024; Green & Malcolm, 2023; Henning et al., 2023; LaFever, 2016; Twyman-Ghoshal & Carkin Lacorazza, 2021). Through a scoping review and an analysis of existing frameworks, we created a crosswalk focused on the aspects of learning the various frameworks addressed (Tucker et al., 2024). Because we are assessment professionals in an academic health center, we had been using a combination of Bloom's Revised Taxonomy (Anderson & Krathwohl, 2001) and Moore's Expanded CME Framework (Moore et al., 2009), which is typically used in continuing medical education and includes the following 7 levels: Participation, Satisfaction, Declarative Knowledge, Procedural Knowledge, Competence, Performance, Patient Health, and Community Health. It became clear that Bloom's is not really a learning outcomes framework, and while Moore's excels at helping us attend to the different levels of learning that can occur in clinical education specifically, it was not a good fit for some schools and programs in our institution such as public health, basic sciences, and health care management.

Our assessment frameworks crosswalk (Tucker et al., 2024), paired with our lived experiences using Bloom's Taxonomy and Moore's CME Framework, helped us determine what was useful from the existing frameworks and what might be missing. From this reflection and analysis, we created the initial *Evidence of Learning and Impact Framework*. The *Evidence Framework* expands beyond Bloom's narrow focus on knowledge and application to consider learning more broadly. From Moore's CME Framework, we adapted the highest levels (Patient Health and Community Health) into our Impact on Others level, making them relevant across disciplines. We also reoriented Moore's lowest levels (Participation and Satisfaction), which felt inconsistent with higher-order outcomes, into equity lenses that can be applied across all levels. This innovation is critical because as faculty plan assessments, they must also examine whether different groups of learners experience inequities, and the framework's design ensures equity is considered early and often. Additionally, our crosswalk with other frameworks revealed overlooked domains such as the "human dimension," "caring," "learning how to learn," "professional identity," and "empathy." We consolidated these into Impact of Learning on Self, capturing personal, relational, emotional, and reflective aspects of learning. Throughout the framework's development, we emphasized assessing "the impact of learners' learning," which is reflected in its final structure. We suggest that the development of this framework is novel because it is a synthesis of many frameworks and is able to be used across many disciplines and contexts.

In addition, one of the more unique features of our *Framework* are the framing questions we developed for each level, outlining concepts both for faculty and for students to consider. These framing questions encompass broader questions such as learners' shifting and growing as they learn, the extent to which they can combine knowledge and skills from different domains, and ways someone who is at the beginning stages of being an expert in their field can think about their potential for impact on others. Generally, we want to encourage thought about the end goals of learning in deep and multifaceted ways, but also present a framework that is streamlined and easy to use.

We found in our scoping review (Tucker et al., 2024) that some existing assessment frameworks are limited in usefulness because they are discipline-specific or only applicable to a specific learning environment which we attempted to avoid. Because the framework was created in the context of work in academic health, we wanted to bring in a group of assessment experts to provide feedback so our framework has wider applicability to multiple higher education settings. For easy reference, Figure 1 shows the graphic of the *Evidence Framework* as we initially conceived it. This paper outlines how we used a Delphi method to iteratively improve our *Framework*.

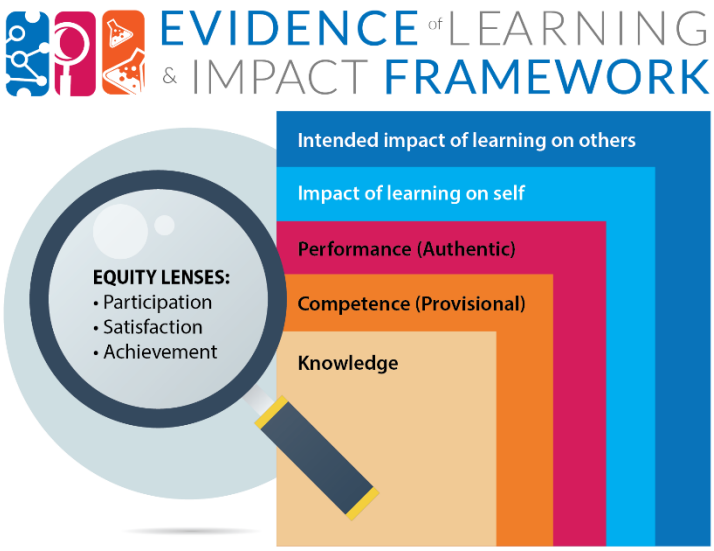


Figure 1. *Initial Version of Evidence of Learning and Impact Framework*

Delphi Approach to Refining Our Framework

Delphi techniques, which are typically used in the technical and natural sciences, are also used in health research to arrive at consensus (Niederberger & Spranger, 2020) and in educational research for a variety of uses including curriculum planning and setting university-wide educational objectives and learning goals (Green, 2014). The assessment experts who participated in our Delphi study provided anonymous feedback using a standardized questionnaire that was adapted over three iterative rounds, allowing us to revise our framework in between each round and reach consensus, which are hallmarks of studies using Delphi techniques (Barrett & Heale, 2020 Diamond et al., 2014; Nasa et al., 2021).

Niederberger and Spranger's 2020 meta-analysis of studies using the Delphi techniques found that it can be difficult to accurately judge the validity of outcomes from a Delphi study because precise details about the relatively complex Delphi technique are not always disclosed. The aim of this article is to share a new higher education learning outcomes framework we have developed while providing more nuance than the norm about the Delphi process we used to refine our framework. Drawing on Niederberger and Spranger (2020), we outline the following key areas for improvement:

1. Clear definition for consensus, along with a discussion of possible factors that may have influenced whether consensus was reached.
2. Inclusion of specific information around the modifications made to the method to help elucidate epistemic objectives and authors' thought processes.
3. Specific characteristics used to determine "who counts" as an expert and the impact of including "lifeworldly" expertise on the Delphi process.
4. Careful thought about the implications of having a small number of experts (in the "low double digits") on the consensus that is reached.
5. Attention to conveying the process for developing and monitoring the questions used in the Delphi surveys which collect experts' thoughts and scores.

We hope that our efforts to share more transparently the decisions we made about our framework revisions will demonstrate the rigor with which our framework was developed and also help higher education colleagues feel more confident in using Delphi techniques themselves.

Delphi Study Team

Before jumping into the details of both our new *Evidence Framework* and the related Delphi study, it is important to understand the study team's positionality. We are three women, two white and one Black, who work in higher education administration in the western United States. We are not health care professionals, but have backgrounds in different aspects of Education (educational psychology, student affairs, K-12, faculty development, and educational linguistics), and are responsible for guiding institutional assessment processes. We have experience in both qualitative and quantitative social science research, but none of us had undertaken a Delphi study before this project. We have strong orientations toward assessment as a tool to shift educational culture/practice and also toward qualitative, social constructivist approaches to thinking about assessment. Among us, we have some previous experience teaching for-credit courses. The institution we work for has publicly and repeatedly set a goal of us working toward becoming an antiracist organization and we have one of the largest groups of Native American students and faculty among academic health centers, though the overall student and employee population is predominantly white.

Preparing for the Delphi Study

The initial tasks we completed before the first round of Delphi are shown in Table 1. A few of these initial tasks merit some more description: determining our criteria for expertise; deciding what our survey questions would be; and our initial attempt at setting a standard for what would count as consensus.

Determining Expertise

Niederberger and Spranger (2020) discuss the ways the concept of expertise has been broadened in the selection of Delphi panels, including placing more value on "lifeworldy" experience (p.3). They cite studies which demonstrate that a more inclusive understanding of expertise leads to more creative and innovative results. In our study, we wanted to be inclusive of the many ways assessment experts gain expertise beyond a formal educational experience, as recommended by Niederberger and Spranger (2020). To that end, in the email we sent out to recruit higher education faculty and staff as

participants we only indicated they needed to be "engaged in assessment in higher education" to participate and did not select for specific titles or roles. Then in the

Table 1. *Tasks Completed Before Delphi Study Began*

Logistics Tasks	Content Tasks
Created list of contacts to invite to participate in the study, including personal contacts and large, international distribution lists related to assessment	Submitted and obtained IRB approval (IRB#: STUDY00024481; found to not be human subject research)
Estimated timeline for Delphi study	Created initial framework graphic in graphic design software
Determined amount and type of incentives for participating, to be distributed ASAP after each round closed	Developed the questions for experts to score, scoring scale, and qualitative, open-ended questions in Qualtrics (Provo, UT) survey
Determined criteria to show participant 'expertise' and developed demographic portion of survey for Delphi participants	Determined how to communicate to the experts about the motivations behind creating a new learning outcomes framework, the background, the framework itself, examples, and framing questions
Decided criteria for reaching 'consensus' and on which question(s) (once we knew what our survey questions would be)	After considering options including recording videos for the experts, we decided to create a slide deck with the narrative directly on the slides along with graphics. Examples of these slides are shared below for each Round

demographics section of the survey itself, we used responses to these questions shown in Table 2 to gauge expertise.

We had only one respondent who indicated they had been working in assessment less than one year. When we looked at the quality of their qualitative feedback in the survey, it was clear they were still novice-level in assessment and their comments lacked useful specificity, so we did not incorporate their feedback nor scores into our analysis. The wording of the second question around expertise was an effort to allow for a more equity-informed way of conceiving of "expertise". We received all "Yes" responses on this question. We received a significant amount of qualitative feedback throughout our Delphi rounds, and the remaining, included participants were able to speak to the questions we asked with insight and helpfully push us where they disagreed. We did not exclude any other participants from the Delphi study. Potential biases in the demographics of our participants are discussed in the Limitations section.

Table 2. *Expertise Questions*

Question	Options
How long have you been engaged in assessment work?	• Less than one year • 1–4 years • 5–9 years • 10+ years
Criteria to give feedback on the Evidence of Learning and Impact framework	• You work/have worked in higher education assessment. • You have credibility with the assessment community through either traditional roles in assessment (such as national and regional organizations, publications, research presentations at national conferences, etc.) and/or non-traditional roles (such as being the "go-to" assessment expert at your institution, have implemented assessment innovations, etc.)
Do you meet these criteria? (Answering "No" will skip you to the end of the survey and no further feedback will be needed.)	• Yes • No

Survey Questions to Collect Expert Feedback

Throughout the three Delphi rounds, we asked our participants to respond to both qualitative and quantitative questions in several categories: reactions to the framework as a whole, reactions to the graphic that represents the framework, and reactions to the individual levels of the framework. Being new to this process, we weren't sure at first if we would use the exact same questions for each round or not, or how much leeway we had to change the questions if we found they weren't working well. We decided to approach the Delphi consensus process with a constructivist research perspective. If our survey questions did not elicit the data we needed, we determined not to hesitate to revise them, of course making sure to note and disclose the details of that change. This allowed us to better hone in on the information we needed each round to further improve our *Evidence Framework*. All survey questions for the three rounds are in Appendices A, B, and C, and you will note that we pivot to some extent in each round to different topics, and also tweaked the wording of some of our quantitative questions.

Initial Consensus Setting

Throughout the Delphi study process, we were working from the assumption that the quantitative Likert scale questions was where we would calculate our official "consensus" to know when we had reached the end and our framework was sufficiently revised, and that the qualitative responses would provide the specific insight needed to make those revisions. Before starting Round 1, we agreed on the following to indicate consensus had been reached: on a Likert scale of 1-6, 90% of respondents would rate the final question for each of five *Evidence Framework* levels, which was "This level requires no editing/updating, or is good as written", as either Strongly Agree or Agree. We also

provided three other statements from the experts to rate about each level which we found very helpful in thinking about consensus: "This level is distinct from the other levels"; "the framing question helped me understand this level"; and "I feel confident I could map assessments appropriately to this level", but we did not initially include those three additional scores for each level in what we thought would be our formal determination of consensus.

The Iterative Delphi Study Rounds

Materials that we sent to the participants each round included a robust slide deck with graphics and explanations of our goals and ways of thinking about our framework, a link to a survey for them to respond to and a list of the survey questions, an information sheet summarizing the study, and to assure participant anonymity, a link to a separate survey to request their gift card incentive. We followed up with one reminder after a week and gave the participants about two weeks to provide feedback. Below, for each of the three rounds we will share examples of the materials we sent out, how many respondents we had, a snapshot of quantitative and qualitative feedback received and status of consensus, insight into what we were thinking as researchers, and a summary of the changes made in response to the experts' feedback.

Round 1

Demographics of Initial Set of Experts

For Round 1, from our recruitment list of contacts and national listservs, we had 39 expert participants provide feedback on our *Evidence Framework*. Additional demographic information on our experts is in Figure 2. Note that our gender data was corrupted, so we excluded it from the information. It is worth noting that only one Delphi participant was affiliated with a health professions institution, and the respondents represented many types of higher education institutions.

Table 3. *Expert Participant Demographics*

Formal Training in Assessment	
Yes	74%
No	26%

Years of Assessment Experience	
1- 4 years	8%
5-9 years	26%
10+ years	66%

Race/Origin	
White	80%
Asian	8%
Prefer not to answer	3%
Hispanic, Latino, or Spanish origin	3%
Mixed	3%
Black or African American	3%

Types of Institutions Represented	Count
Community, Junior, and Technical Colleges	13
Faith-Based Colleges	8
For-Profit Institution	4
Four-Year Colleges and Universities	39
Health Professions Institutions	1
International Colleges	1
Liberal Arts College/University	14
Minority-Serving Institutions (tribal colleges, HBCU, HSI, etc.)	8
Special Focus Institution (arts/design/music, military, law, etc.)	2

Age Range	
Under 30	5%
30-49	54%
50-69	41%
70 and above	0%

Materials Reviewed in Round 1

The slide deck the experts read through for Round 1 included study objectives and timeline; background information about why we created the Evidence of Learning and Impact Framework; an introduction to the new framework; practical examples of how it could be used; an explanation of the equity lenses and how we reframed those from Moore's Satisfaction and Participation levels; and references and resources.

The following two slides formed the primary material we were looking for feedback on. You can see in Figure 3 that we had adjusted the coloring of the Framework (but none of the wording) for the initial round. In Figure 4 we provided one example and 1-2 framing questions for each of the framework levels. You can also see the way we communicated with our experts in the wording on the slides. Appendix A contains the complete set of questions our experts responded to in Round 1 after reviewing the slides.

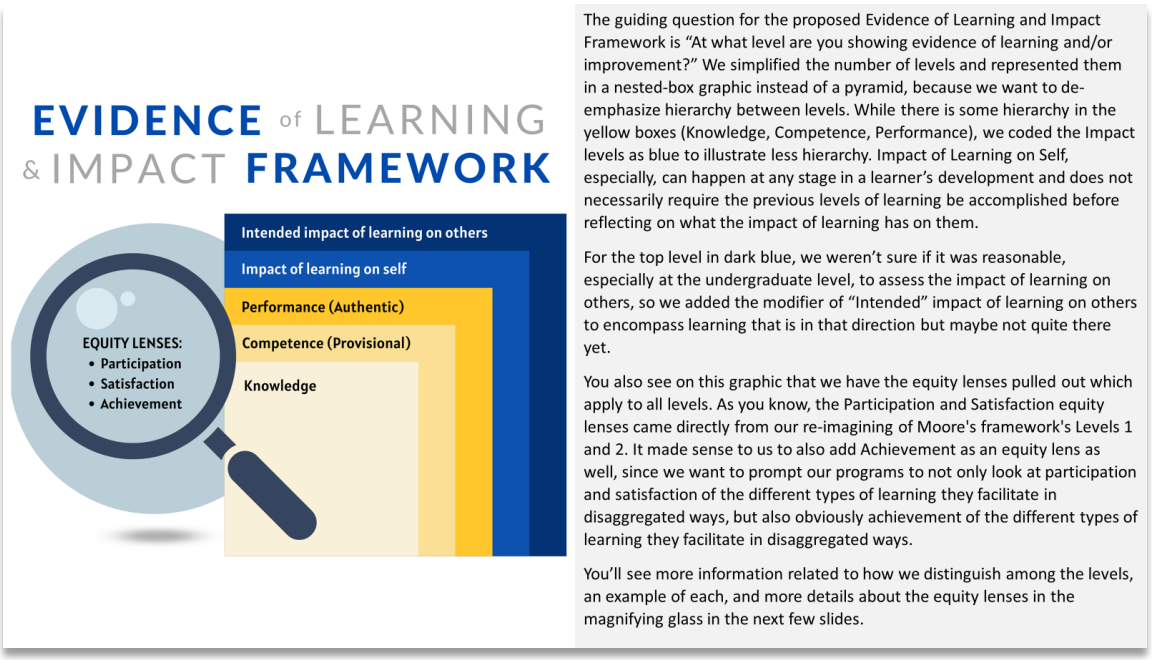


Figure 3. *Round 1 Evidence Framework Graphic*

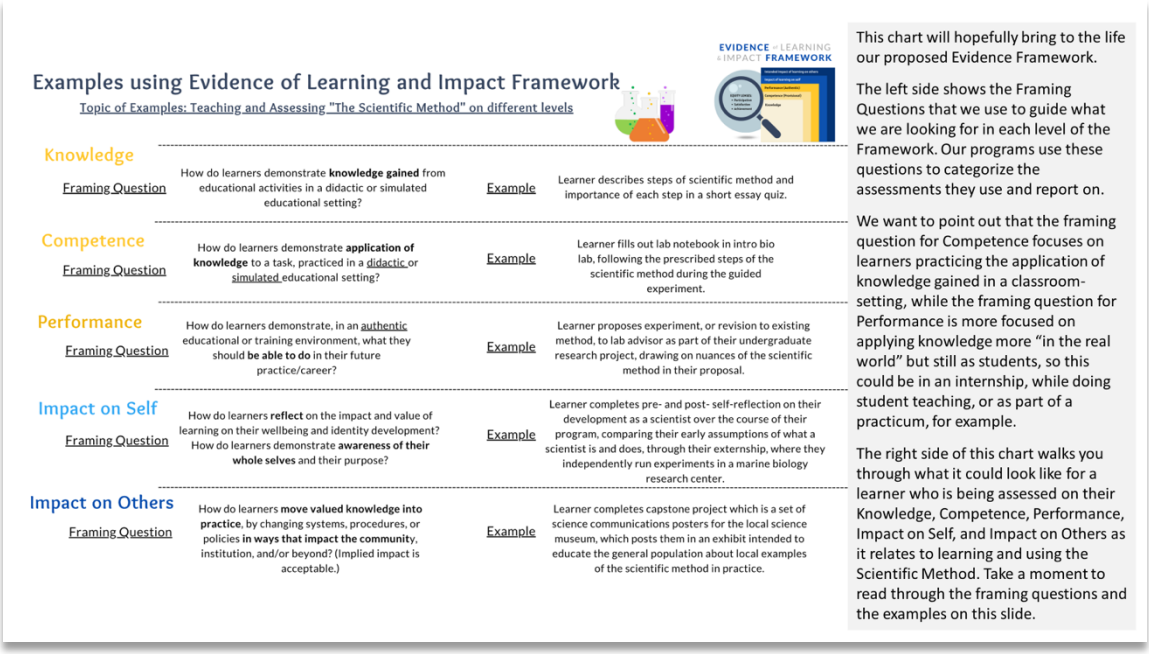


Figure 4. *Round 1 Examples and Framing Questions*

Table 4. *Quantitative Results from Round 1*

Question Text	Response Count (N)	High Agreement (%) (Strongly Agree, Agree)	Unclear (%) (Somewhat Agree, Somewhat Disagree)	Low Agreement (%) (Disagree, Strongly Disagree)
Framework as a Whole				
This framework would be easy to apply to a variety of disciplines.	39	89.74%	10.26%	0.00%
This framework would be easy to apply to interprofessional education (defined as learning with, from, and about each other)	39	94.87%	5.13%	0.00%
This framework makes space for diverse ways of knowing.	39	89.74%	10.26%	0.00%
This framework centers the learner in assessment (assessment for learning vs. of learning)	39	79.49%	20.51%	0.00%
Knowledge				
This level is distinct from the other levels.	35	85.71%	8.57%	5.71%
The framing question helped me understand this level.	35	80.00%	20.00%	0.00%
I feel confident I could map assessments appropriately to this level.	35	85.71%	11.43%	2.86%
This level requires no editing/updating, or is good as written.	35	48.57%	45.71%	5.71%
Competence (Provisional)				
This level is distinct from the other levels.	35	77.14%	17.14%	5.71%
The framing question helped me understand this level.	35	85.71%	11.43%	2.86%
I feel confident I could map assessments appropriately to this level.	35	80.00%	11.43%	8.57%
This level requires no editing/updating, or is good as written.	35	51.43%	31.43%	17.14%
Performance (Authentic)				
This level is distinct from the other levels.	35	74.29%	22.86%	2.86%
The framing question helped me understand this level.	35	77.14%	20.00%	2.86%
I feel confident I could map assessments appropriately to this level.	35	62.86%	34.29%	2.86%
This level requires no editing/updating, or is good as written.	35	31.43%	51.43%	17.14%
Impact of Learning on Self				
This level is distinct from the other levels.	35	91.43%	2.86%	5.71%
The framing question helped me understand this level.	35	71.43%	22.86%	5.71%
I feel confident I could map assessments appropriately to this level.	35	77.14%	14.29%	8.57%
This level requires no editing/updating, or is good as written.	35	42.86%	34.29%	22.86%
Intended Impact of Learning on Others				
This level is distinct from the other levels.	35	91.43%	5.71%	2.86%
The framing question helped me understand this level.	35	77.14%	20.00%	2.86%
I feel confident I could map assessments appropriately to this level.	35	57.14%	40.00%	2.86%
This level requires no editing/updating, or is good as written.	35	34.29%	45.71%	20.00%

After Round 1, three of our overall Likert-scale questions had achieved consensus: "This framework would be easy to apply to interprofessional education", "This framework would be easy to apply to a variety of disciplines", and "This framework makes space for diverse ways of knowing". Instead of asking those questions again in Round 2, we added two new overall questions to further dig into improving the framework: "The *Evidence Framework* graphic is an effective way to visualize the framework" and "Do you agree with the decision to move participation and satisfaction assessment methods out of the framework in order to use them as equity lenses?"

Qualitative Results from Round 1

Two of the authors took the qualitative input from the experts and, using Word, organized the data into "representative quotes about strengths" for each level, "quotes about weaknesses/possible action items", and "recurring themes" for each level. We paid attention to the quotes about strengths because wanted to make sure that we knew when to protect portions of our *Framework* from significant changes. Because we were primarily looking for specific action items that came out of the data that we could consider and use to make concrete changes to the framework, the examples, and the framing questions in a short timeframe, we did not more formally code the data in software. We used the Word doc as a comprehensive potential to-do list which we organized thematically and which framed our discussions as we met over the next couple of weeks to prepare for Round 2 and we documented our decisions about the feedback in the Word document itself to be sure we were considering all of the comments and on the same page as a research team. During these conversations we discussed the contradictions we saw in the experts' responses, the ways they aligned, and tried to reconcile these tensions in our revisions. We re-thought the boundaries between levels, the wording of framing questions and examples, and documented rationale behind our changes to incorporate into the next Round's slide deck. In total for the Delphi study we analyzed approximately 45 single-spaced pages of comments from our experts, which led to a substantial understanding of our participants' perspectives on our work.

During Round 1, qualitative data on the Knowledge level indicated that Knowledge was similar to other frameworks and fairly easily understood, basic, foundational, and clear. However, we needed to include more examples from other disciplines. We were also encouraged to provide a more encompassing use of "knowledge" and determine whether or not skills were included in Knowledge or in another level. The phrase "knowledge of, about, how, and why", which we added to the framework graphic after this round, came directly from one of our expert participants. Some other frameworks separate knowledge into smaller buckets, distinguishing among declarative and procedural knowledge, for example, or knowledge you remember vs. knowledge you apply. We decided we wanted faculty to put more focus on the other levels and not worry so much about precisely which kind of knowledge they are fostering. Knowledge of has to do with general awareness of a topic, knowledge about has to do with cataloging facts, knowledge how is remembering the steps to doing something, and knowledge why speaks to understanding the context, when different knowledge is applicable, and impacts of using that knowledge. This isn't meant to be thorough conceptualization of knowledge, and there may be other parts of knowledge that align to this level as well.

Qualitative data on the Competence (Provisional) level indicated there was confusion regarding terms like "provisional", "didactic", "simulated", and not everyone was clear on how Competence was different from Performance.

> "Depending on the content or task, I can imagine it being difficult to distinguish between Competence and Performance levels. Both require application of knowledge to a task."

Qualitative data on the Performance (Authentic) level indicated the experts had mixed feelings about whether Performance is useful or too close to Competence as noted above. The experts also expressed that it might be hard for some (perhaps less experiential-education focused) disciplines to identify what is Performance in their realm.

Qualitative data on the Impact of Learning on Self level indicated fairly strong support for incorporating a focus on the impact of learning on the learner in our framework. We also heard that the experts didn't understand how well-being fit into this level, that the examples need to be altered to focus on the student and not the task, and that we should expand our framing questions to be broader.

> "I wish Impact on Self more explicitly recognized our students as whole people and more explicitly referenced a growth mindset in learning."

Qualitative data on the Intended Learning of Impact on Others level indicated mixed feedback on the use of the word "Intended" in the name of the level, and perhaps too much overlap among this level, Performance (Authentic), and Impact on Self.

> "Very few learners will ever reach Impact on Others with all content they learn… especially not in bachelor's level programs."

Qualitative data on the graphic itself led us to revise it for Round 2.

> "For the graphic: Less emphasis on hierarchy… thus more of a perception that the learning taking place at all levels is valued."

Insight into Authors' Conversations

As we read through the qualitative data, we discussed whether faculty of any individual course should feel compelled to align with all five levels of the *Evidence Framework*, which was an assumption that some of the experts' feedback seemed to make. We decided that, theoretically, any program should be able to align to all five levels, but that an instructor of a specific course should feel free to align to whichever ones made sense for their content and course goals.

We also decided that we needed to provide more examples and framing questions in Round 2 for experts to review and hopefully get a better understanding of what we meant by each level as we honed in on the differences among them. And we wanted to provide more explanation and examples related to our Equity Lenses in Round 2 because we realized we did not receive specific enough feedback on that part of our framework

from our first Delphi round. Part of this conversation was us also determining that there were perhaps more components we needed feedback on than we had initially realized.

Consensus After Round 1

A slide (Figure 4) was shared in Round 2's slide deck and summarizes the state of consensus and timeline.

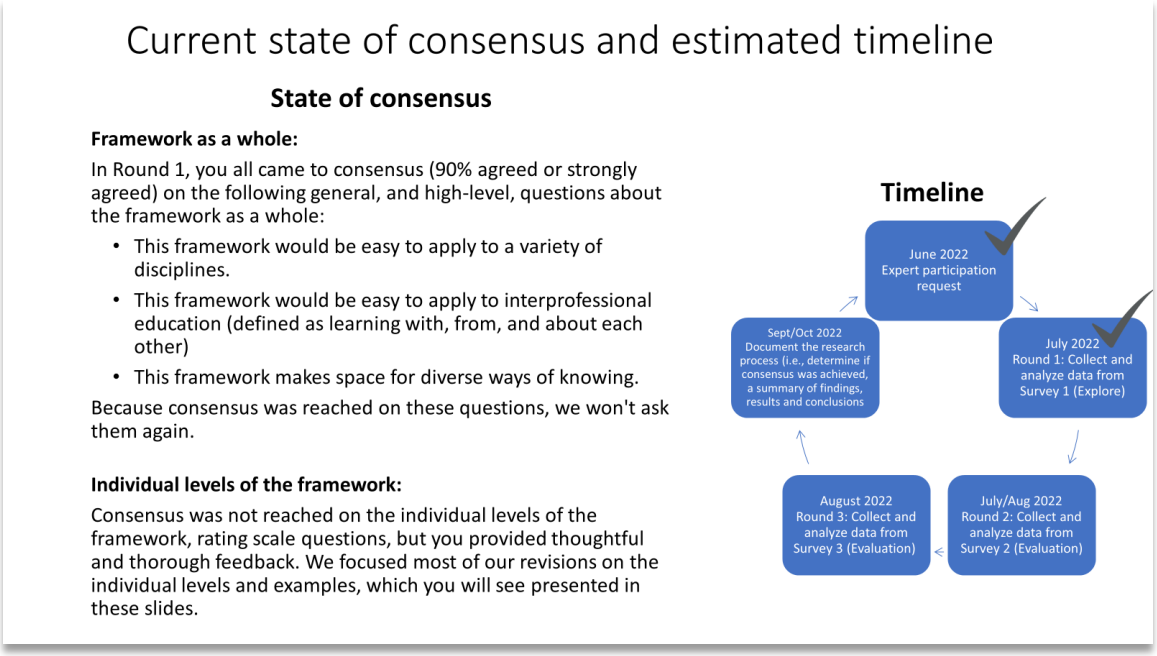


Figure 4. *State of Consensus after Round 1*

Round 2

Expert Participants Round 2

For Round 2, from the original set of 39 experts, 28 (72%) of them provided feedback on the revised *Evidence Framework* materials. No new experts were added to Round 2.

Materials Reviewed in Round 2

The slide deck the experts read through included a reminder of the crosswalk and other influences we were basing our framework on: a summary of the main changes we made, more in-depth exploration of the Equity Lenses, and for each individual level, a pair of slides that outlined additional framing questions, examples across the disciplines, and potential aligned assessment approaches. Examples of these are provided below.

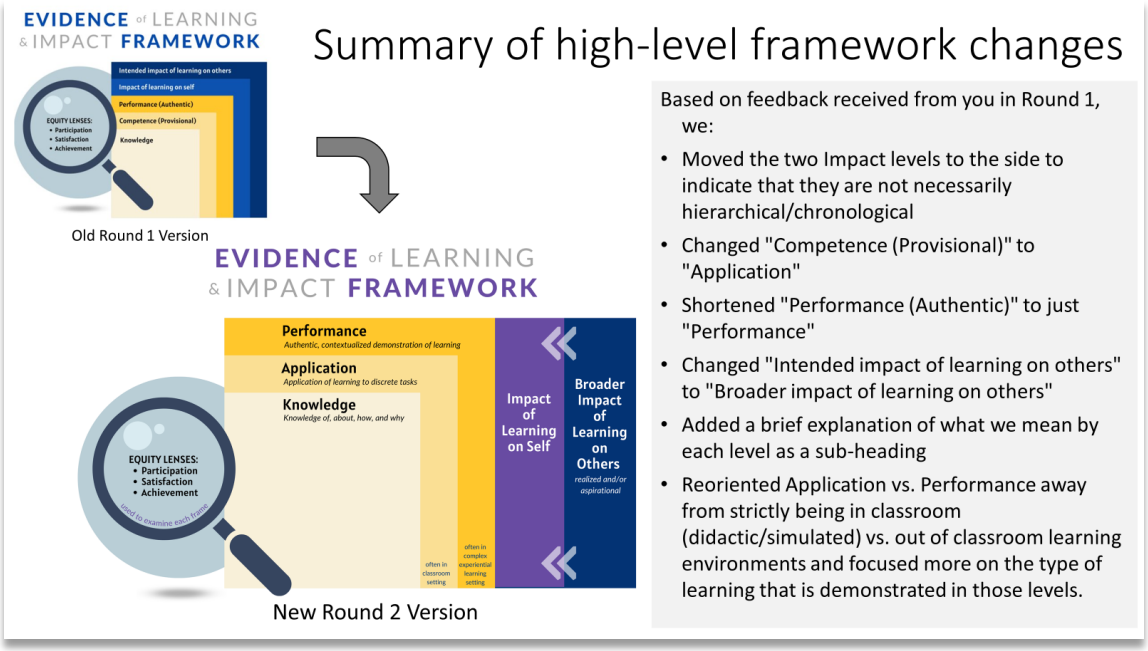
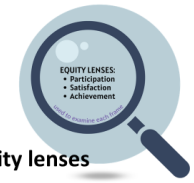


Figure 5. *Changes made after Round 1 to the Evidence Framework graphic*

Experts indicated in Round 1 that we needed to provide more clarity around why we moved Participation and Satisfaction out of the main structure of our previous (Moore's Outcomes Levels) framework, added Achievement, and re-imagined them all as lenses through which we examine equity. As we were creating the crosswalk (Tucker, Jacobs, and Moreno, 2024), we realized Participation and Satisfaction aren't evidence that learning is occurring, but rather they tell us who is there and whether they feel welcome/comfortable in the learning setting. The following two slides (Figures 6 and 7) were used to show how we defined Participation, Satisfaction, and Achievement Equity Lenses, the process by which our programs demonstrate the Equity Lenses portion of the framework, and a good example of what one program at our institution submitted for the equity lens question.

Finally, in the Round 2 materials, based on the qualitative feedback we received in Round 1, we provided some additional framing for reviewers to keep in mind when reading the set of slides on the individual levels (see figure 8).

How we use these lenses in our yearly assessment planning and reporting process



Reporting – Closing the loop equity question

Tell us about **one assessment activity** the program is **analyzing with an equity lens**. This example should:

- Identify an assessment activity that is **ripe for improvement** using and equity lens,
- Describe an approach/data source used to analyze equity to improve learning (e.g. **disaggregation** of who participates and/or how satisfied they are, achievement, etc.), and
- Describe how the program is using the data to **inform decision making.**

Examples of how to analyze equity lenses:

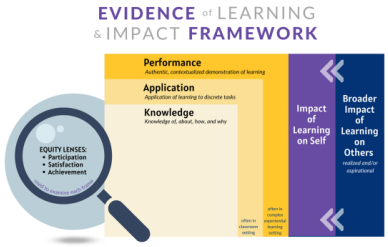
- **Participation**: What demographic patterns in data exist (e.g., comparing attendance and demographic data)?
- **Satisfaction**: How are learners experiencing the learning environment? How are learning outcomes affected by the learning environment?
- **Achievement**: Who is excelling and who is not? What grades are which students earning and is there an equity gap?

How programs use the equity lenses

- A program maps their planned assessment activities to the five Evidence Framework levels and plans yearly targets for success.
- Then, at end of the yearly cycle, the program looks at their data, reflects on the question at left and decides what they can do as educators to improve learning as a whole and to examine any differences they see in participation, satisfaction, and/or achievement.
- When gaps in participation, satisfaction, or achievement are found that differ by demographics or other means, educators will gather additional data from students, staff, and other stakeholders as needed to clarify the issue. This will help them determine if curricular or programmatic changes need to be made and what those should look like to provide a more equitable learning environment and/or experience.
- After changes are made, the program reports on the previous cycle and tells a story of how they used an equity lens to uncover a gap and how they are addressing it

Figure 6. *Round 2 Equity Lenses*

Examples of Closing the Loop with Equity Lenses from Our Programs



Achievement Lens, Performance and Impact on Self Levels:
Note: Preceptor is a faculty mentor in clinical learning

From the end of term, one-on-one student meetings and preceptor evaluations and feedback, we noticed that our students who identify as Latino/a/e scored themselves lower on all self-evaluations and received lower scores from preceptors on “leadership”, “initiative”, and “self-advocacy”. As a result, we are investigating how we can best support students' self-efficacy in leadership roles. We are working with preceptors to provide them with the knowledge and skills to engage and empower these students to feel comfortable stepping into a leadership position, as well as, unconscious bias awareness training. At our end-of-term meetings, we are providing an opportunity for ethnically diverse students to provide us with feedback to continue to refine our approach.

Figure 7. *Round 2 Example of Equity Lens in Use*

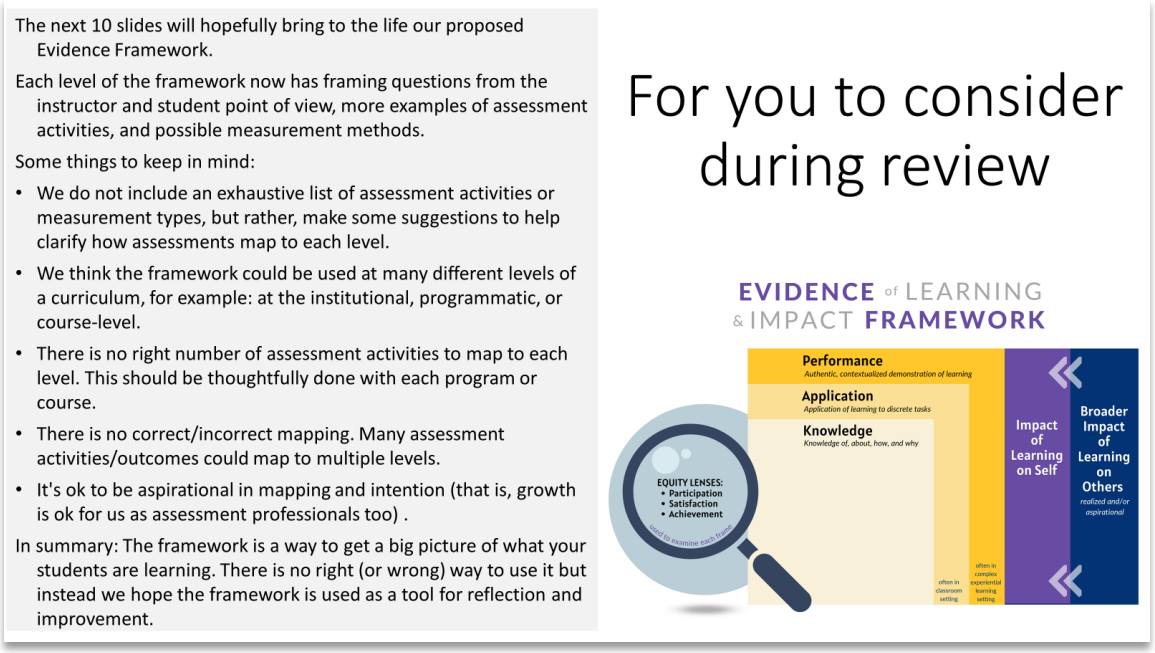


Figure 8. *Round 2 To Keep in Mind When Scoring Individual Levels*

Figure 9 and 10 illustrate two slides for the Performance Level that are representative of the kind of information we shared on each level during Round 2. You can see we shared changes made to the label of the level, revised framing questions both for students and for educators, and provided more examples. Appendix B contains the complete set of questions our experts responded to in Round 2.

Quantitative Results from Round 2

After Round 2, our overall Likert-scale question "This framework centers the learner in assessment" had achieved consensus (at least 90% Agree or Strongly Agree) and almost 93% of the participants responded "Yes" that they agreed with moving Participation and Satisfaction into Equity Lenses. For the specific levels, significant progress was made in getting closer to consensus on the Knowledge, Impact on Self, and Impact on Others levels. We also calculated whether we had reached consensus if we were to expand it to include Slightly Agree and it did not make much difference, so we decided to keep consensus as at least 90% that chose Agree or Strongly Agree. You can see from Table 5 that some of the individual questions about the levels were starting to reach consensus after Round 2.

Performance*

*This level was formerly called "Performance (Authentic)". Based on feedback, and in the interest in clarity, we have renamed it Performance.

Framing question(s) for educators:

- *To what extent do learners integrate their learning and demonstrate their depth and breadth of knowledge and skills as they negotiate a complex task?*

Framing question(s) for students:

- *How am I showing I am drawing on my knowledge and skills to navigate complex tasks?*
- *How am I adapting to new or conflicting information as I apply my knowledge to real-life tasks?*
- *How am I using what I've learned to make decisions based on multiple inputs and with various audiences and goals in mind?*
- *How do I unite my technical skill and creative expression in an authentic setting?*

Potential aligned assessment approaches:

- Comprehensive/Holistic Review: Program Committee/Panel Review - TAC/DAC reviews, theses, dissertations, capstones, portfolio, proposal defense
- External Review of Student Work(s): Manuscript feedback, external national assessments, IRB approval, grant review, peer reviewed blogs and presentations
- Internal Performance Observation: Simulation/Clinical Exercise, Research lab notebooks, portfolio, dissertation/thesis/capstone, peer evaluation
- Self-Assessment: Written, oral or other, portfolio development

Figure 9. *Round 2 Performance Framing Questions*

Performance - Examples

- ***Public Health***: Students draft a policy memo relevant to their internship audience, which is then evaluated by their preceptor and/or other internship stakeholders.
- ***Computer Science***: Students work collaboratively in their internship placements to develop a decision tree.
- ***Applied Linguistics:*** During student teaching, discuss the cognitive benefits of bilingualism with a parent who is upset that their Kindergartener still isn't reading in either language.
- ***English Literature***: Student publishes a poem they wrote for class in a local literary magazine, navigating revisions to the poem and professional expectations from the publisher, with coaching from their professor.
- ***Student Life*:** Student leads a team-building activity in the group they started and the student life trainer gives feedback.
- ***Career Services***: Student makes decisions about which student employment position and which clubs to participate in based on identified values.
- ***Undergrad science:*** Learner notices undergraduate research project is not resulting in usable data and proposes a revision to the existing method to lab advisor, drawing on knowledge of the scientific method and ethical research practices.
- ***Clinical***: In a simulated or real clinical setting, students take a history and physical exam on a patient, and then create a differential diagnosis.

Figure 10. *Round 2 Performance Examples*

Table 5. *Quantitative Results from Round 2*

Question text	Response Count (N)	High Agreement (%) (Strongly Agree, Agree)	Unclear (%) (Somewhat Agree, Somewhat Disagree)	Low Agreement (%) (Disagree, Strongly Disagree)
Framework as a Whole				
This framework centers the learner in assessment	28	92.86%	7.14%	0.00%
The Evidence Framework graphic is an effective way to visualize the framework	28	85.71%	14.29%	0.00%
This level is distinct from the other levels.	28	96.43%	0.00%	3.57%
Knowledge				
The framing questions helped me understand this level.	28	82.14%	17.86%	0.00%
I feel confident I could map assessments appropriately to this level.	28	96.43%	3.57%	0.00%
This level requires no editing/updating, or is good as written.	28	57.14%	35.71%	7.14%
Application				
This level is distinct from the other levels.	28	82.14%	14.29%	3.57%
The framing questions helped me understand this level.	28	71.43%	25.00%	3.57%
I feel confident I could map assessments appropriately to this level.	28	85.71%	14.29%	0.00%
This level requires no editing/updating, or is good as written.	28	42.86%	46.43%	10.71%
Performance				
This level is distinct from the other levels.	28	64.29%	17.86%	17.86%
The framing questions helped me understand this level.	28	71.43%	17.86%	10.71%
I feel confident I could map assessments appropriately to this level.	28	67.86%	14.29%	17.86%
This level requires no editing/updating, or is good as written.	28	46.43%	25.00%	28.57%
Impact of Learning on Self				
This level is distinct from the other levels.	27	100.00%	0.00%	0.00%
The framing questions helped me understand this level.	28	85.71%	7.14%	7.14%
I feel confident I could map assessments appropriately to this level.	28	78.57%	17.86%	3.57%
This level requires no editing/updating, or is good as written.	28	64.29%	25.00%	10.71%
Broader Impact of Learning on Others				
This level is distinct from the other levels.	28	92.86%	7.14%	0.00%
The framing questions helped me understand this level.	28	75.00%	21.43%	3.57%
I feel confident I could map assessments appropriately to this level.	28	78.57%	17.86%	3.57%
This level requires no editing/updating, or is good as written.	28	60.71%	32.14%	7.14%

Qualitative Results from Round 2

We organized our qualitative data again in Word to determine what changes we wanted to make based on expert feedback. Here are the major takeaways we found at each of the framework levels:

Qualitative data showed there was some confusion as to whether demonstration of skills belonged at the Knowledge or Application level of the framework.

Qualitative data on the Application level indicated the experts like the choice of "Application" for this level more than "Competence". Much of the substantive feedback was on needing to further distinguish between Application and Performance, namely that the location of the assessment (in class vs. in experiential learning) is not perfectly mapped onto more discrete tasks vs. more complex ones. A couple quotes from participants demonstrate the kind of feedback we received on this topic:

> "I have questions about the knowledge and skills part of the explanation… where do skills become Application or Performance vs. Knowledge"

> "As much as you can, play up the complex or multi-faceted/varied circumstances or environment of Performance to differentiate it from the Application level."

Qualitative data on the Performance level indicated both that Performance is too much of a leap from Application and also that they aren't distinguishable enough. Clearly,

the experts were struggling with the way we presented two levels in the slide deck as well. We did get a few really positive comments about Performance that made us think we are on the right path with it:

> "Robust framing questions and potential aligned assessment approaches are very clear and useful; All of the framing, aligned assessments, and examples are really strong. I think this is my favorite level as it is currently written; The scaling up to complexity, authenticity, and real-life settings and tasks; It represents complex thinking/learning/doing for a variety of contexts and circumstances."

> "The holistic aspect of it invites a long-term, wide lens of examination to assess student progress, which is unique from the other two levels. I see Knowledge and Application as really course-based, held within specific assignments, while Performance is more program-based, with several possible points of data."

Qualitative data on the Impact of Learning on Self level was primarily centered on the scope and wording of the framing questions we provided. Experts pointed out that the thrust of the Impact on Self framing questions for educators weren't fully aligned with the framing questions for students, for example. We also received some positive comments about this level:

> "This is the part of learning that is often left out of assessment frameworks, so I am glad to see it called out here."

> "I love this level! We do not often ask students to reflect on how their learning is affecting them."

> "Productively student-centered and metacognitive."

Qualitative data on the Broader Impact of Learning on Others level indicated the experts were not sure they liked the addition of "Broader" to the title, but they didn't agree on using "Intended" either. Much of the rest of the feedback was on finer points of the wording in the framing questions. A few positive comments about this level were also shared:

> "It helps higher education to reestablish itself as a public good by centering the emphasis on building healthy communities as the grand goal of learning."

> "I like that it covers the recognition, awareness, and intentionality of learning in relation to others. the part about recognizing new ways to perceive the world around them is particularly compelling."

Insight into Authors' Conversations

Much of our discussion after Round 2 was centered on trying to decide what we really meant by Application vs. Performance. We firmly believe that there is something different about a nursing student *performing* learning such as drawing blood in front of a family, while a baby is crying, while trying to speak a second language with patients versus *applying* learning to doing the blood draw on a mannequin arm in a simulation center. Fundamentally, we decided it wasn't just about where the learning or assessment was occurring, but also about the requirement to draw on lots of disparate sets of knowledge to accomplish the task that distinguished Performance from Application. However, we know this is not a perfect distinction.

We also discussed whether we wanted to call the different parts of the framework "levels" or some other term (category, group, set, kind, measure, section, frame, type). We decided to add a question about this to Round 3 to see what the experts thought.

Consensus After Round 2

A slide (Figure 11) was shared in Round 2's slide deck and summarizes the state of consensus and timeline.

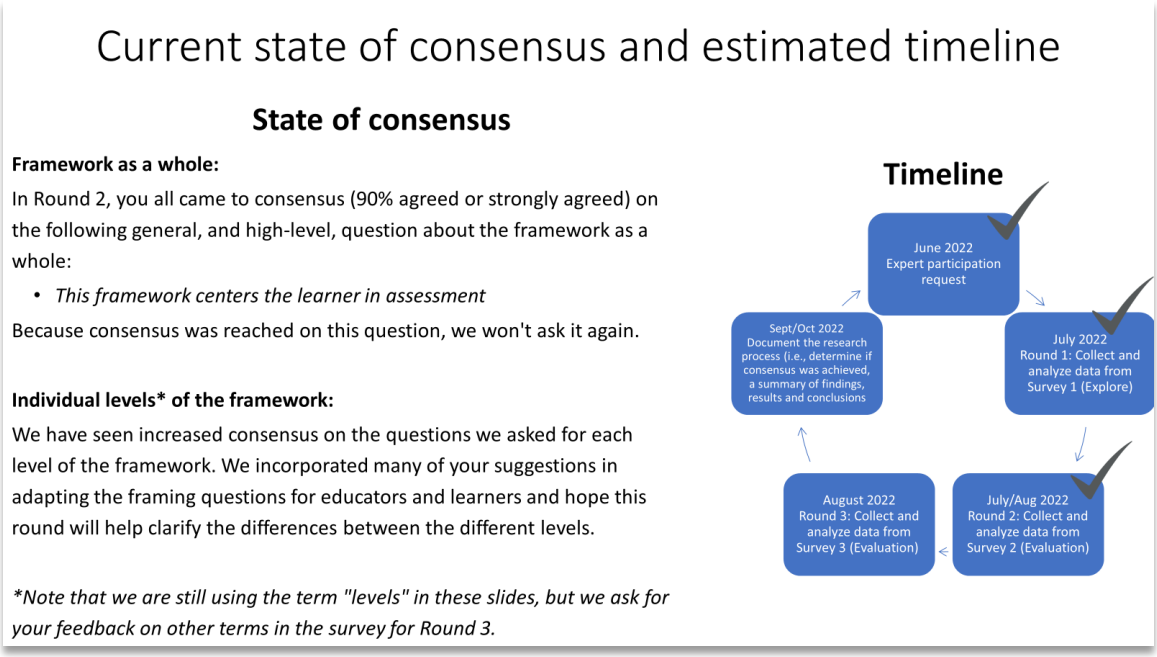


Figure 11. *State of Consensus after Round 2*

Round 3

Expert Participants Round 3

From the original set of 39 experts, we had 22 Round 3 expert respondents (22 participants is 78% of those who responded in Round 2, or 56% of those who responded in Round 1). No new experts were added to Round 3. We estimate that it took at least an hour for the experts to respond to each round of the Delphi study, which makes the numbers dropping each round unsurprising, and also consistent with literature (Boel et al., 2021). We think we were able to retain the number of experts that we did because we provided a relatively quick turnaround from round to round, we were responsive to their feedback, and we shared with them transparently what changes we made and why.

Niederberger and Spranger (2020) identified an area described for improvement we highlighted in the introduction to this paper: '4. Careful thought about the implications of having a small number of experts (in the low double digits) on the consensus that is reached.' Our Delta rounds, from 22-38 expert participants each, generated a lot of rich, insightful data in line with what you would expect from a rigorous qualitative study. If we had received the same amount of detailed feedback from 80 or 100 experts we would have been drowning in data. And because we maintained a relatively high bar in terms of reaching consensus on the quantitative data, we are confident that the revisions we made to the Evidence Framework were well-conceived.

Materials Reviewed in Round 3

The slide deck for Round 3 focused primarily on clarifying and differentiating the five levels of the framework and finalizing the framing questions. See figure 12 for a summary of changes made after round 2. From qualitative feedback in Round 2, we decided that the framing questions were a more powerful way to define the different levels, and pivoted to focus just on those during Round 3 instead of on refining our examples. The focus on framing questions is communicated in Figure 13. Figure 14 is an example of the revisions we made to each level's slide focused on framing questions. Performance is shown for continuity with the example provided above in Round 2 (see figures 9 and 10). Figure 15 shows how we conceived the differentiation between Application and Performance after two rounds of feedback telling us we needed to make that clearer. Appendix C contains the complete set of questions our experts responded to in Round 3.

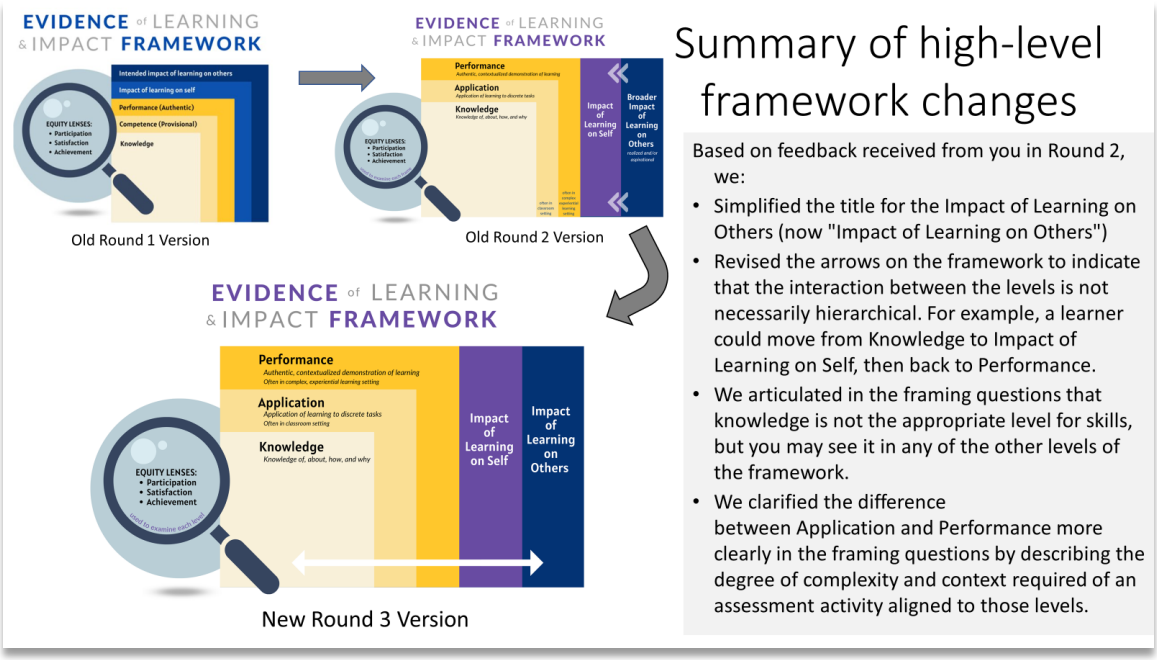


Figure 12. *High-Level Changes Made after Round 2 to the Evidence Framework*

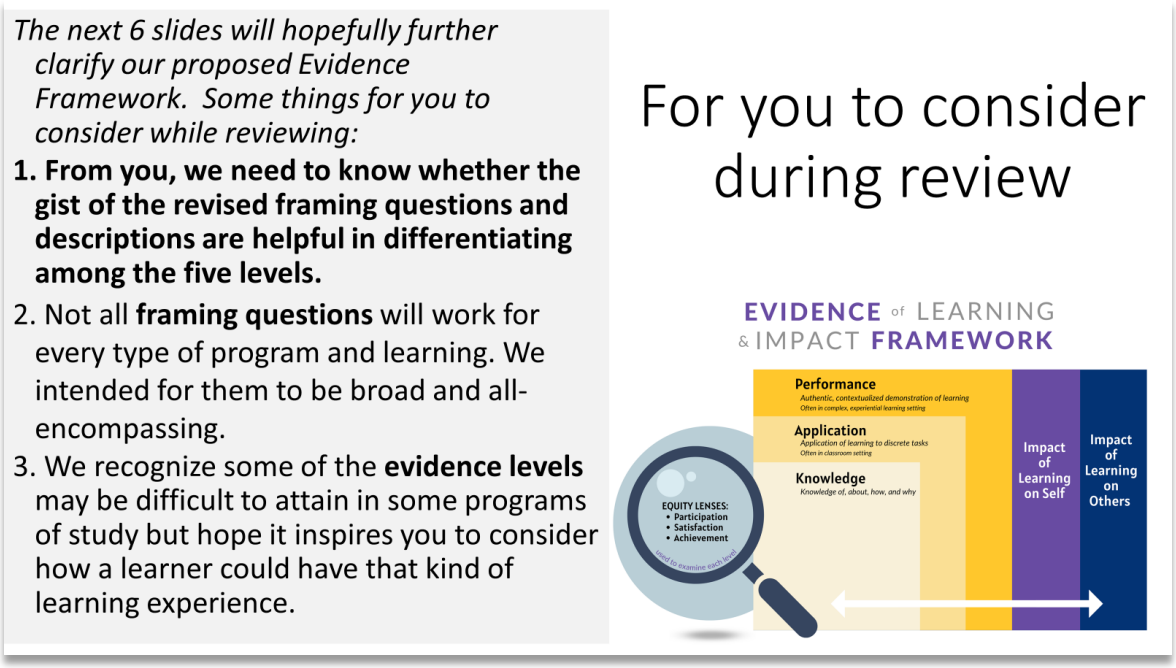
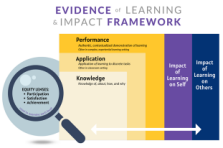


Figure 13. *To Consider While Reviewing Round 3*

Performance

Authentic, contextualized demonstration of learning; often in complex or novel contexts. Usually seen in experiential learning settings that require navigating the situation as a whole by using knowledge/skills from different domains.



Framing question(s) for educators:

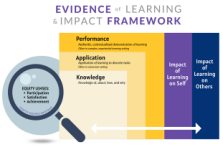
- To what extent do learners integrate their learning and demonstrate their depth and breadth of knowledge and skills as they navigate complex tasks and/or situations?
- To what extent do learners navigate unexpected challenges as they apply their learning? Do they transfer knowledge in messy situations?

Framing question(s) for students:

- How do I show that I am able to thoughtfully apply knowledge and skills in complex ways under real constraints and conditions? Am I able to move beyond the script or formula? How am I adapting to new or conflicting information?
- How do I draw on knowledge and skills from disparate subject areas?
- How am I using what I've learned to make decisions based on multiple inputs and with various audiences and goals in mind?
- How do I integrate technical skill with creative expression to demonstrate learning and accomplish objectives?

Figure 14. *Round 3 Performance Description and Framing Questions*

Application versus Performance



Application is:

Defined by application of learning to discrete tasks. This often occurs in a classroom setting or in controlled contexts and may rely on formulas or scripts. Learners practicing skills would fall under application.

Performance is:

Defined by an authentic, contextualized demonstration of learning. It often occurs in complex or novel contexts and is usually seen in experiential learning settings that require navigating the situation as a whole by using knowledge/skills from different domains.

Application Examples	Performance Examples
Computer Science: Using sample user interface data from a popular software application, students work with classmates to draft a list of suggested updates/changes to improve user experience with software.	**Computer Science**: Student identify a user interface design problem within the company where they are interning, research possible solutions, identify appropriate data to collect, and propose a solution to improve user experience. (This same assessment activity could be mapped to Impact on Others, if the project were actually completed, and user experience data was gathered and analyzed for further improvement).
English Literature: Students practice writing poems that focus on elements such as repetition, symbolism, and figurative language.	**English Literature:** Student publishes a poem they wrote for class in a local literary magazine, navigating revisions to the poem and professional expectations from the publisher, with coaching from their professor.
Undergrad Science: Learner follows steps of scientific method during intro bio lab and writes up experiment in lab notebook.	**Undergrad Science:** Learner notices undergraduate research project is not resulting in usable data and proposes a revision to the existing method to lab advisor, drawing on knowledge of the scientific method and ethical research practices.

Figure 15. *Round 3 Application vs. Performance*

Quantitative Results from Round 3

After Round 3, we determined that we had reached consensus, though not exactly in the way we had originally identified when we started the Delphi study. In looking at Round 2 data, we worried that wording our main consensus question for each question "This level requires no editing/updating, or is good as written", and requiring Strongly Agree or Agree at 90% was perhaps setting a rather perfectionistic bar for ourselves. We decided to slightly revise the wording of that question in Round 3 to "This level requires minor editing/updating, or is good as written." We still wanted to keep consensus at 90% Strongly Agree or Agree.

Table 6. *Quantitative Results from Round 3*

Question Text	Response Count (N)	High Agreement (%) (Strongly Agree, Agree)	Unclear (%) (Somewhat Agree, Somewhat Disagree)	Low Agreement (%) (Disagree, Strongly Disagree)
Knowledge				
This level is distinct from the other levels.	22	100.00%	0.00%	0.00%
The framing questions helped me understand this level.	22	95.45%	4.55%	0.00%
I feel confident I could map assessments appropriately to this level.	22	95.45%	4.55%	0.00%
This level requires minor editing/updating, or is good as written.	22	72.73%	9.09%	18.18%
Application				
This level is distinct from the other levels.	22	90.91%	9.09%	0.00%
The framing questions helped me understand this level.	22	100.00%	0.00%	0.00%
I feel confident I could map assessments appropriately to this level.	22	100.00%	0.00%	0.00%
This level requires minor editing/updating, or is good as written.	22	72.73%	13.64%	13.64%
Performance				
This level is distinct from the other levels.	22	90.91%	9.09%	0.00%
The framing questions helped me understand this level.	22	95.45%	4.55%	0.00%
I feel confident I could map assessments appropriately to this level.	22	100.00%	0.00%	0.00%
This level requires minor editing/updating, or is good as written.	22	77.27%	4.55%	18.18%
Impact of Learning on Self				
This level is distinct from the other levels.	22	100.00%	0.00%	0.00%
The framing questions helped me understand this level.	22	100.00%	0.00%	0.00%
I feel confident I could map assessments appropriately to this level.	22	90.91%	9.09%	0.00%
This level requires minor editing/updating, or is good as written.	22	81.82%	9.09%	9.09%
Impact of Learning on Others				
This level is distinct from the other levels.	22	100.00%	0.00%	0.00%
The framing questions helped me understand this level.	22	95.45%	4.55%	0.00%
I feel confident I could map assessments appropriately to this level.	22	90.91%	9.09%	0.00%
This level requires minor editing/updating, or is good as written.	22	86.36%	0.00%	13.64%
Framework as a Whole				
The Evidence Framework graphic is an effective way to visualize the framework	22	100.00%	0.00%	0.00%
Assuming minor edits based on feedback, the evidence framework, as a whole, works for me.	22	95.45%	4.55%	0.00%

Round 3 data showed that with our new, slightly modified question, in combination with all the revision made to the *Evidence Framework*, that we now had between 72% and 86% of experts Agreeing or Strongly Agreeing that each level was done. However, when we looked at the three sub-questions for each level ("this level is distinct", "framing questions help me understand this level", and "I could map to this level confidently"), each one of those had reached consensus of at least 90%. We decided to declare consensus reached in this way and ended the Delphi study.

Additional quantitative data collected in Round 3 was whether we should use a different term than "level" when discussing the *Evidence Framework*. 52% of the experts said we should keep "level", and the term with the second most votes (33%) was "category". We chose to retain the use of "level" for now.

Qualitative Results from Round 3

In Round 3, feedback we received included wording tweaks for the Application framing questions; that we had more adequately addressed the difference between Application and Performance; and that we needed to reconsider the use of "scholarship" in our framing of the Impact of Learning on Others. We received many comments similar to "Solid, clear to understand", "You nailed it", "Framing questions are fantastic", and some really lovely comments at the end of the survey.

> "THANK YOU! Your courage and practice in the development of this framework are so impressive. As I apply your very framework to the work you have done, I think you check all of the boxes. I'd like to be able to use some of your resources in my work conducting faculty development on my campus, and I hope you consider licensing this for others to use with an attribution. Thank you."

> "This has been a great process and the framework is better because of it. Thank you for your contribution to the field."

An analysis of the qualitative data also showed that the degree of participant confusion had decreased, as did the scope of the changes that were being suggested as we progressed through the three rounds, which helped to confirm the consensus calculations that came out of the quantitative data.

Summary of Final Changes Made

We integrated many of the Round 3 suggestions we received into the final version of the framing questions and added some additional text to the graphic.

Discussion

Final Version of Evidence of Learning and Impact Framework

The final version of the *Evidence of Learning and Impact Framework* is shown in Figure 16. We invite faculty, programs, schools, and universities to consider adopting and, importantly, adapting it with attribution for their own use.

We hope that the *Evidence of Learning and Impact Framework* will challenge educators to think differently, more broadly, and more deeply about the kinds of learning they foster and assess. We also hope the framework pushes higher education, and in our case, especially STEM, toward a space where multiple narratives and multiple interpretations of information are valued, and where students are given opportunity to reflect meaningfully on the holistic impact of their learning on themselves and those around them.

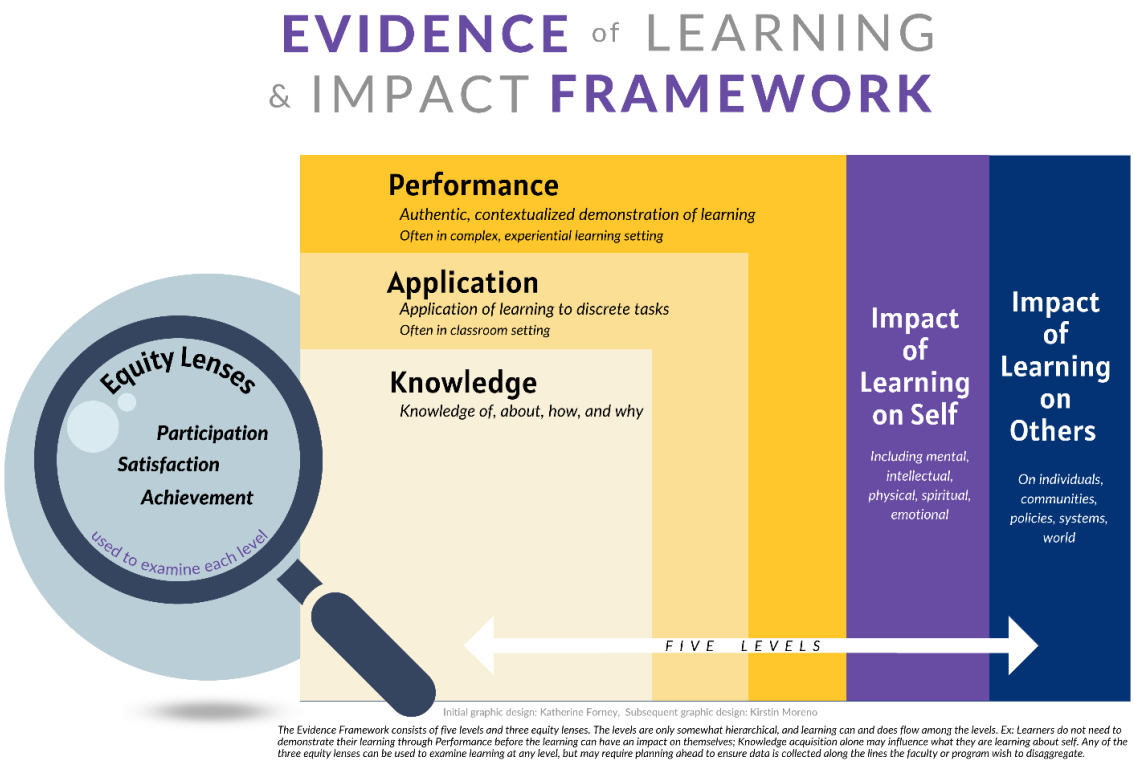


Figure 16. *Final Version of Evidence of Learning and Impact Framework (or "Evidence Framework")*

Implementation Challenges

We have been using our *Framework* since 2020 and have noted a few implementation challenges. At a high level, it can be challenging to shift a group of faculty from a known paradigm such as Bloom's Taxonomy to a new way of thinking about assessment. We re-visited the distinctions between Bloom's, Moore's and our *Evidence Framework* with assessment staff and faculty over time to help them fully understand the aims behind the new framework. In addition, a couple of the levels are a bit challenging to learn how to assess or measure. For example, incorporating Impact of Learning on Self has not been as difficult as we expected, possibly because reflective practice is common in the health professions, but Impact on Others has taken more work to orient to. Most of our programs have maybe only one or two Impact on Others-aligned activities, as this level is a big lift and may apply well to capstone projects. To support faculty in learning a new framework, it was important to reassure them that they don't need to master every aspect at once and acknowledging this has been helpful. Finally, the boundaries between Application and Performance, and also between Performance and Impact on Others can be a bit of a grey area, but we have found that looking at the intentions of the assessment activity and the method used to assess the learning helps to clarify the correct level to assign. In some cases, it's appropriate to assign more than one level, so we've built in flexibility to allow for that. Overall, we've found that the *Evidence of Learning and Impact Framework* aligns well with the needs of our academic and student affairs departments, and we've encountered fewer barriers compared to the previous frameworks we used.

Additional Resources to Assist with Implementation

To support readers in effectively applying the *Evidence of Learning and Impact Framework*, we have included a set of supplementary resources. Specifically, the final list of framing questions—revised in response to feedback from Round 3—can be found in Appendix D. These questions are intended to facilitate immediate implementation of the framework, particularly for those who require minimal additional guidance.

In addition, Appendix E contains a list of theoretical example activities that align to the *Evidence Framework* from various disciplines that was revised based on feedback from Round 2 of the Delphi study. The most significant revisions we made based on the Delphi study to these example activities in Appendix E are:

- The addition of History examples to capture different aspects of Impact on Self and how it can be connected to the Knowledge level and to affect.
- The addition of phrasing to the Performance level examples to emphasize the importance of bringing together multidisciplinary knowledge and skills, based on feedback such as "As much as you can, play up the "complex" or multi-faceted/varied circumstances or environment of this learning demonstration to differentiate it from the "Application" level."

Considerations for Adopting the Evidence of Learning and Impact Framework

Based on our experience having faculty and staff implement the Evidence Framework at OHSU, we recommend the following:

1. Begin by having programs align their assessments to the framework levels, supporting this work with charts, trainings, and/or consultation.
2. Consider allowing programs to map to both an older framework (e.g. Bloom's Taxonomy, etc.) and the Evidence Framework at the same time in order to scaffold understanding about the Evidence Framework.
3. Build in checks on quality alignment to Evidence levels, such as trained peer reviewers, giving feedback and suggesting realignment/changes where needed.
4. Programs often struggle most with the two Impact levels, so we recommend additional training and exercises to reinforce the types of measurement which work best for these two levels (e.g. journals and self-knowledge tests for Impact on Self, or community impact data for Impact on Others). These two levels draw on practices from Impact Evaluation, so we recommend you familiarize yourself.
5. Once initial alignment and instruction has happened, introduce the equity lenses and how they might be applied in context, for example, by asking learners their perspectives on the learning environment, or by disaggregating student outcome performance data and examining the results for possible inequities.
6. Starting small is key, whatever that looks like at your institution. We used pilot groups, double (or triple) mapping, and trained framework champions to help with implementation.

Please adapt the framework levels and equity lenses to fit your institutional context.

Limitations

Our Delphi study and subsequent *Evidence Framework* would be more robust if we had been able to recruit more expert participants from places outside of North America, especially because we imagine conceptualizations around ideas of Impact on Self and Impact on Others may be quite different in cultures that are more collectivist in nature, or that perhaps have less centering of whiteness, colonization, capitalism, and similar orientations. We have mitigated this somewhat by ensuring that the scoping review (Tucker, Jacobs, and Moreno, 2024) we conducted included research and thinking about learning and assessment from countries in the global south or indigenous perspectives. We drew on these studies to help us construct some of our framing questions for each level and to refine our equity lenses, but more geographically dispersed Delphi experts would have been useful partners in this work.

Future Work

In future work, we hope to share more details about how this framework could be used (in assessment planning and reporting at the institutional level, in program- and course-level assessment practices), more examples of the equity lenses in action, and a discussion of assessment methods that pair well with the framework levels. The final list of framing questions, revised based on Round 3 feedback, are in Appendix D so the framework can be implemented now for those who need less guidance.

Additional future work could examine whether the Evidence of Learning and Impact Framework is indeed applicable to different sectors of higher education including career technical education and the humanities given that we have only been using it in an academic health center. Future scholarship could also expand upon specific ways to approach the equity lenses in an assessment framework of this scope. Finally, the Impact on Self and Impact on Others levels of the Evidence Framework are especially rich spaces to think differently and more broadly about the impact of learning on learners and those around them. We would love to see more assessment scholarship that dives into these levels.

Using Delphi Techniques for Improving Educational Innovations:

Delphi techniques used in this study were a very powerful way to significantly improve the rigor of the framework we had already put a lot of time and thought into developing. As this was the first time we employed a Delphi process, we did not know what to expect in terms of the quality or amount of feedback we would receive, or whether our participants would choose to continue to stick with the study. Happily, the experts provided an abundance of input, very generously sharing their thoughts with us, and prompted many nuanced conversations as we figured out how to synthesize their feedback and be responsive to their insights. The Delphi process forced us to really think about the different components of our framework and what we needed to address with each. We fully endorse employing Delphi techniques if you are looking to refine something that has broad appeal and have the time to put in the work to run a Delphi study well. We did not realize how long it would take to pull together all of the materials for each round, to clean the data to complete the consensus calculations, and to sort through the qualitative feedback. Estimate more time than you think!

There are various things about Delphi techniques that are not intuitive. Those include determining exactly how to determine consensus, for example: which questions(s), at which level of agreement, from which percentage of respondents, and whether you should average scores in doing so, etc. Determining whether the qualitative feedback plays a role in consensus-determination is also not something that occurred to us intuitively. On these things you'll need to trust your gut in addition to looking at how other Delphi studies have handled determining consensus, and then just be transparent about what you did (Niederberger and Spranger, 2020). There are many ways to approach running a Delphi study.

Tips and Crucial Mindsets for Using Delphi Techniques

- Anonymity is key. A small portion of the (anonymous) feedback the experts had to give us was quite critical, but necessary for us to hear.
- Take time to really determine what the different, concrete aspects of your project are that you want expert feedback on. Determine if you will ask for that all at once in the first round, or if you will space it out, prioritizing different aspects for different rounds of the Delphi process.
- It is critical to carefully document your process throughout so you can provide detailed transparency in your own scholarship that draws on Delphi techniques.
- Be transparently flexible with consensus setting. Check that you're not being perfectionistic, but maintain high standards. Determining when to claim consensus is an art, not a science.
- Be responsive and take input you receive in good faith, and share back what you changed based on that input. It really makes a difference to the experts who take time to participate in the study.

Appendices

Appendix A: Round 1 Survey Questions
Appendix B: Round 2 Survey Questions
Appendix C: Round 3 Survey Questions
Appendix D: Framing Questions for the *Evidence of Learning and Impact Framework*
Appendix E: Evidence of Learning and Impact Framework Example Activities

Acknowledgements: The authors wish to thank the OHSU Assessment Council for their collaboration on initial versions of the Evidence Framework; our Delphi Study participants who truly provided an abundance of useful and insightful feedback as we went through our three rounds; and Kathie Forney, for her initial graphic design of our Evidence Framework.

Previous Presentations: Some details from our Delphi study and revised Evidence of Learning and Impact Framework were presented by the authors at the International Association for Medical Education (AMEE) in Lyon, France, Aug 27-31, 2022, at the IUPUI Assessment Institute in Indianapolis, Indiana, Oct 9-11, 2022, and at AALHE, Jun 3-6, 2024 in Portland, Oregon.

References

Alegría, M., Thurston, I. B., Cheng, M., Herrera, C., Markle, S. L., O'Malley, I. S., Porter, D., Estrada, R., & Giraldo-Santiago, N. (2024). A learning assessment to increase diversity in academic health sciences. *JAMA Health Forum*, *5*(2), e235412. https://doi.org/10.1001/jamahealthforum.2023.5412

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Allyn & Bacon. Boston, MA. ISBN: 978-0801319037

Barrett, D., & Heale, R. (2020). What are Delphi Studies? *Evidence-Based Nursing*, *23*(3), 68–69. https://doi.org/10.1136/ebnurs-2020-103303

Boel, A., Navarro-Compán, V., Lawendé, R., & van der Heijde, D. (2021). Two different invitation approaches for consecutive rounds of a Delphi survey led to comparable final outcome. *Journal of Clinical Epidemiology, 129*, p. 31–39. https://doi.org/10.1016/j.jclinepi.2020.09.034

Green, R. A. (2014). The Delphi technique in educational research. *SAGE Open, 4*(2). https://doi.org/10.1177/2158244014529773

Green, M., & Malcolm, C. (2023). Degrees of change: the promise of anti-racist assessment. *Frontiers in Sociology, 8*. https://doi.org/10.3389/fsoc.2023.972036

Henning, G., Rice, A., Heiser, C., & Lundquist, A. (2023). Equity-centered assessment practices: Survey findings and recommendations. *Research & Practice in Assessment, 18*(2), p. 20–30. https://www.rpajournal.com/dev/wp-content/uploads/2024/03/Equity-centered-Assessment-Practices-RPA.pdf

LaFever, M. (2016). Switching from Bloom to the Medicine Wheel: Creating learning outcomes that support Indigenous ways of knowing in post-secondary education. *Intercultural Education*, *27*(5), 409–424. https://doi.org/10.1080/14675986.2016.1240496

Moore, D. E., Jr, Green, J. S., & Gallis, H. A. (2009). Achieving desired results and improved outcomes: integrating planning and assessment throughout learning activities. *The Journal of Continuing Education in the Health Professions*, *29*(1), 1–15. https://doi.org/10.1002/chp.20001

Nasa, P., Jain, R., & Juneja, D. (2021). Delphi methodology in healthcare research: How to decide its appropriateness. *World Journal of Methodology*, *11*(4), 116–129. https://doi.org/10.5662/wjm.v11.i4.116

Niederberger, M., & Spranger, J. (2020). Delphi technique in health sciences: A map. *Frontiers in Public Health*, *8*, 457. https://doi.org/10.3389/fpubh.2020.00457

Tucker, C., Jacobs, S., & Moreno, K. (2024). Branches from the same tree: A scoping review of learning outcomes frameworks. *Intersection: A Journal at the Intersection of Assessment and Learning*, *5*(4), 53–82. https://doi.org/10.61669/001c.123960

Twyman-Ghoshal, A., & Carkin Lacorazza, D. (2021, March 31). Strategies for antiracist and decolonized teaching. *Faculty Focus*. https://www.facultyfocus.com/articles/equality-inclusion-and-diversity/strategies-for-antiracist-and-decolonized-teaching/ retrieved April 30, 2024.