DR ERIN E SHORTLIDGE (Orcid ID : 0000-0001-8753-1178)

Article type : Academic Practice in Ecology and Evolution

Submission type: Academic Practice in Ecology and Evolution

Title: *A Resource for Understanding and Evaluating Outcomes of Undergraduate Field Experiences*

Authors: Erin E. Shortlidge[1*], Alison Jolley[2], Stephanie Shaulskiy[3], Emily Geraghty Ward[4], Christopher N. Lorentz[5], Kari O'Connell[6]

Running title: *Assessment of Undergraduate Field Experiences*

Addresses:
[1*]Portland State University, Department of Biology, Portland, OR 97201
[2]University of Waikato, Te Puna Ako – Centre for Tertiary Teaching & Learning, Hamilton, New Zealand 3240
[3]University of Michigan, Biological Station, Ann Arbor, MI 48109
[4]Rocky Mountain College, Geology Program, Billings, MT 59102
[5]Thomas More University, Department of Biological Sciences, Crestview Hills, KY 41017
[6]Oregon State University, STEM Research Center, Corvallis, OR 97331

***Author for Correspondence:**
eshortlidge@pdx.edu; Portland State University, PO Box 750, Portland OR 97201; 503-725-9305

Keywords: field experiences, undergraduates, assessment, learning outcomes, inclusion

ABSTRACT

Undergraduate field experiences (UFEs) are a prominent element of science education across many disciplines; however, empirical data regarding the outcomes are often limited. UFEs are unique in that they take place in a field setting, are often interdisciplinary, and include diverse students. UFEs range from courses, to field trips, to residential research experiences, and thereby have the potential to yield a plethora of outcomes for undergraduate participants. The UFE community has expressed a strong interest in better understanding how to assess the outcomes of UFEs. In response, we developed a guide for practitioners to use when assessing their UFE using an evidence-based, systematic and iterative approach. This essay guides practitioners through the steps of: identifying intended UFE outcomes, considering contextual factors, determining an explicit assessment approach, and using the information to inform next steps. We provide a table of common learning outcomes and potential assessment tools, vignettes to illustrate using the strategy, and suggestions for practical application of the strategy. We aim to support comprehensive and aligned assessment of UFEs, leading to more inclusive and reflective design, and ultimately improved student outcomes. We urge practitioners to move towards evidence-based advocacy for continued support of UFEs.

INTRODUCTION

Background

Conducting research, collecting data, and teaching students outside of a laboratory or classroom setting is commonplace across disciplines. For many scientists, being “in the field” is paramount to the work that they do (Wilson 1982, Cutter 1993, Rudwick 1996). Therefore, in numerous disciplines, engaging undergraduates in authentic field experiences or experiences that take place in the field (**undergraduate field experiences, UFEs**) is not only expected and intuitive (Dressen 2002), but considered central to training goals (Gold et al. 1994, Fleischner et al. 2017, Giles et al. 2020). For the purposes of this paper, we borrow from the work of colleagues (Fleischner et al. 2017, Morales et al. 2020, O’Connell et al. 2021) to define what we are considering to be a UFE**.** UFEs are designed explicitly with student learning in mind and occur in a field setting where students engage with the natural world, or through a virtual experience, meant to mimic an experience in the field. UFEs can take place in a variety of settings and durations including immersive, residential courses or programs at field stations and marine labs, short field trips as part of traditional on-campus university courses, or long, multi-day field trips. The COVID-19 pandemic has further encouraged the development of remote UFEs, and challenged us to reflect on how lessons in field educational design might apply beyond in-person settings (e.g., Barton, 2020). The discussion that follows mostly applies to in-person and remote UFEs. Further, we are not limiting our discussion of UFEs to field biology, geoscience, or natural history courses, as we are aware of the wide-range of disciplines with UFEs, and aim to be inclusive of these experiences.

Some have argued that a student's undergraduate experience in disciplines such as biology, ecology, and the geosciences is not complete without a UFE (Cutter 1993, Nairn 1999, Petcovic et al. 2014, Klemow et al. 2019). A survey of participants at the Geological Society of America meetings (2010 & 2011), showed that the majority (89%) of survey participants felt that field experiences were vital to geoscience education, and that the bulk of the value lies in cognitive gains, and to a lesser degree, sustained interest in the field (Petcovic et al. 2014). The Governing Board of the Ecological Society of America showed strong support of UFEs by including field work and the ability to apply natural history approaches as two of the ecology practices in the recently adopted Four-Dimensional Ecology Education Framework (Klemow et al. 2019).

Participating in a UFE can spark students' interest in the scientific topic being explored in the field (Dayton and Sala 2001, LaDue and Pacheco 2013, Petcovic et al. 2014), increase student cognitive gains in disciplinary content (Easton and Gilburn 2012, Scott et al. 2012), improve student understanding of the process of science (Patrick 2010), foster development of discipline-specific technical skills (Peasland et al. 2019) and increase persistence in STEM fields (Jelks and Crain 2020). UFEs can also have far-reaching impacts, even changing the trajectory of students' lives by influencing career choices, or solidifying long-term commitments to the environment (Palmer and Suggate 1996, Barker et al. 2002). UFEs have been identified as critical contributors to students' development of a sense of place (Semken 2005, Billick and Price 2010, Van Der Hoeven Kraft et al. 2011, Semken et al. 2017, Jolley et al. 2018a) as well as fostering a resonance with Indigenous peoples and Traditional Ecological Knowledge (Cajete 2000, Riggs 2005).

Despite these key outcomes, some have voiced fears about field experiences going "extinct," and have sounded alarm bells for stakeholders to consider how to gain further support for such experiences (Barker et al. 2002, Whitmeyer et al. 2009a, Swing et al. 2021). There is a widespread occurrence of, and in many cases, fervent advocacy for undergraduates learning in the field, yet given their prevalence, there is a lack of systematically collected data on specific outcomes resulting from the diversity of possible field experiences (Mogk and Goodwin 2012). Practitioners (field instructors, directors, coordinators and staff) want to understand the efficacy of their individual programs, while universities and funding agencies require evidence of success for continued support of undergraduate field programs. Stakeholders across disciplines have made it clear that more empirical studies that test claims of positive student outcomes are needed for continued support of UFEs (Smith 2004, Clift and Brady 2005, NRC 2014, O'Connell et al. 2018). This is particularly true as it relates to improving equity, access, and inclusion in the field (NRC 2003, Brewer and Smith 2011, Wieman 2012, Morales et al. 2020). Collecting evidence of student outcomes will help to identify opportunities and challenges for supporting the inclusion of all students in UFEs, and aid in tackling some of the challenges with inclusion that we already know exist in UFEs (O'Connell et al. 2021).

Practitioners report a strong interest in collecting evidence of outcomes from their UFEs for iterative improvement, to demonstrate value of their programs, and to contribute to broader understanding of field learning, but do not feel confident in their ability to measure student outcomes, given that it is not their expertise (O'Connell et al. 2020). Indeed, most of the studies that have measured outcomes from UFEs are conducted by education researchers, trained in quantitative and/or qualitative research methods. To meet practitioners where they are, and support mindful, efficacious assessment of UFEs, we: **1)** present a resource for practitioners to use when they want to assess UFE outcomes and improve their programs and courses, **2)** address how assessment and evaluation of UFE outcomes can help practitioners better design inclusive field experiences, and **3)** identify an existing pool of instruments that align with intended student outcomes of UFEs.

Conceptualization of this Paper

The authors of this paper are members and founders of the Undergraduate Field Experiences Research Network (UFERN; www.ufern.net), a NSF-funded Research Coordination Network focused on fostering effective UFEs. UFERN brings together diverse perspectives and expertise to examine the potentially distinctive learning and personal growth that happens for students when they engage in UFEs across the range of disciplines and formats. During a UFERN meeting (2019), it became apparent that undergraduate field educators from across disciplines were frequently requesting help in how to collect empirical evidence about complex student outcomes from UFEs (O'Connell et al. 2020). The work presented here emerged from conversations at that UFERN meeting and is a collaboration between STEM education researchers, social scientists, and undergraduate field educators from multiple disciplines, to directly address calls for guidance on assessing UFEs.

Suggested Strategy for Assessing UFEs

We advocate that stakeholders work to understand and evaluate their UFEs or UFE programs in clear alignment with the unique goals of each individual field experience. Reflecting best practices in designing learning environments that support student gains, we draw from the process described as 'backwards design' (Wiggins et al. 1998). Importantly, this method emphasizes the alignment of UFE design to the outcomes being measured. We build from a 'how to' guide designed for assessing course-based undergraduate research experiences (CUREs) presented by Shortlidge and Brownell (2016) and have expanded and tailored the guide to be specific to UFEs. Figure 1 is to be used as a guide and as a mechanism for reflection, allowing practitioners to refine a UFE to better serve the students, meet the intended outcomes, and/or change and build upon data collection methods already in place.

We aim to provide a guide that is inclusive to those who intend to assess, evaluate, and/or conduct education research on UFEs, and therefore will describe how these are separate but interrelated and likely overlapping actions. In order to clarify potential misunderstandings, we explain the language that we use regarding assessment, evaluation, and research.

We use the word ***assessment*** when we are referring to measuring student learning outcomes from UFEs. Assessment tools refer to the instruments that are used to collect the outcome data (e.g. a survey, rubric, or essay). Assessments can use qualitative (e.g. interviews), quantitative (e.g. surveys), or a mix of approaches (Creswell 2013).

A ***programmatic evaluation*** might aim to holistically understand the experience that all or individual stakeholders have in a UFE; the evaluation could include students, instructors, program directors, community partners, etc. To evaluate something is to determine its merit, value or significance (Patton 2008), and program evaluation has been described as "the systematic assessment of the operation and/or outcomes of a program or policy, compared to a set of explicit or implicit standards as a means of contributing to the improvement of the program or policy" (Shackman 2008). Thus, an evaluation of a UFE would determine the appropriate assessment methodology and identify if programmatic goals are being met. Such information can inform how a UFE can be improved. Evaluation is often conducted by an external evaluator who may work with the UFE leadership team to develop a plan, often through the creation and use of a site-specific logic model (Taylor-Powell and Henert 2008). An evaluation can target a range of UFEs, from a singular disciplinary program, or an entire field station's season of hosted UFEs.

The collection of empirical evidence about a UFE, which can be gathered through assessment and evaluation, and adds new knowledge, could potentially be used for education ***research***. Authors Towne & Shavelson state that: *"…education research serves two related purposes: to add to fundamental understanding of education-related phenomena and events, and to inform practical decision making… both require researchers to have a keen understanding of educational practice and policy, and both can ultimately lead to improvements in practice."* (Towne and Shavelson 2002, p. 83).

If the aim is to publish research outcomes from a UFE, practitioners will likely need to submit a proposal to an Institutional Review Board (IRB). The IRB can then determine if a human subjects' research exemption or expedition protocol will be necessary. If an IRB protocol is needed, this should occur *before* data collection begins. Gaining IRB approval is contingent on researchers having been certified in human subjects' research and a robust and detailed research plan that follows human subjects' research guidelines. Thus, conducting education research on UFEs requires advance planning, and ideally would be conducted in partnership with or with advisement from education researchers. Typically, if a study is IRB approved, participants of the study need to consent to their information to be used for research purposes.

Publishing outcomes may be desirable, but not all data will be collected in a way that yields publishable results, yet those results may be highly informative to practitioners and UFE programs. Designing effective formative assessments to understand and modify a UFE might be the most appropriate workflow before engaging in intentional research studies on the outcomes of a UFE. Importantly, we do not advocate that one method is better, or more or less appropriate than another; the approach should depend on the aims and intentions of the stakeholders and the resources available.

Guide to Assessing UFEs and Sample Vignettes

Fig. 1 is presented as a guide for practitioners to use for understanding the outcomes of a UFE. The green arrows signify that each box informs the other, and iterative reflection and refinement are a key aspect of informed evaluation and assessment. The guide includes four key components: **I)** Identifying the intended student and/or programmatic **outcomes** for the UFE; **II)** Considering the **context** of the UFE, which may include any number of factors related to: setting, duration, timing, discipline, student identity, and accessibility of the UFE; **III)** Defining an assessment **approach** that is appropriate for the context and in alignment with the intended outcomes; **IV)** Utilizing the outcomes and approach to inform and refine **next steps** in the UFE.

To highlight diverse UFEs and give realistic examples of assessment and evaluation approaches, we present four examples of UFEs, referred to as ‘vignettes’ (Fig. 2). The vignettes provide examples of how one can apply the components of the guide (Fig. 1) to a given UFE, and at the end of the paper we present two of the vignettes in a more detailed narrative, offering examples that synthesize the ideas presented (*Expanded Vignettes*).

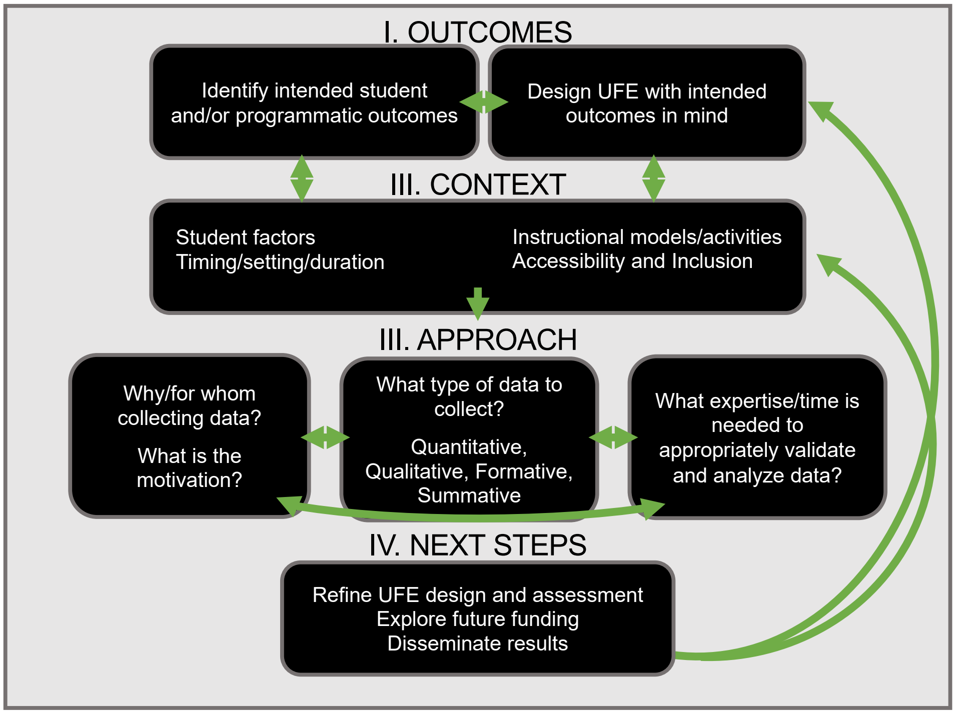


Figure 1. Guide for Assessing Undergraduate Field Experiences (UFEs). The figure presents a guide to walk practitioners through assessing their UFE. The green arrows signify that each box informs the other, and iterative reflection and refinement are a key aspect of informed evaluation and assessment.

I. Identify the Intended Outcomes From the UFE

The main focus of this work is to provide the tools and resources needed such that stakeholders can confidently assess if students are meeting expected learning outcomes from UFEs (e.g. students expand their knowledge of endemic amphibians; students report an increased interest in environmental sustainability efforts); however, programmatic outcomes and goals (e.g. participants are involved in community engagement and scientific knowledge-building activities) are also critical components of this type of learning environment, and thus are also represented in example vignettes (**Fig. 2**).

We draw upon Bloom's Taxonomy of Learning (Bloom and Krathwohl 1966, Anderson et al. 2001) to aid practitioners in considering the possible outcomes from UFEs. The taxonomy describes three fundamental domains of learning: the cognitive, affective, and psychomotor domains. Studies about UFEs demonstrate that students may experience outcomes across all of these domains and more (Boyle et al. 2007, Stokes and Boyle 2009, Scott et al. 2012, Petcovic et al. 2014, Scott et al. 2019, O'Connell et al. 2020). Cognitive outcomes from a UFE could include: an improved ability to explain plant species interactions, accurately identify geological formations, or solve a problem using an interdisciplinary lens (Fuller et al. 2006, Bauerle and Park 2012, Tripp et al. 2020). Affective outcomes could include: a newfound interest in a subject, such as conservation; motivation to continue seeking out field learning experiences; or, development of a connection to place (Boyle et al. 2007, Simm and Marvell 2015, Jolley et al. 2018a, Scott et al. 2019). Outcomes in the psychomotor domain could include: the improved ability to geolocate, collect and measure sediment in a lake with the appropriate instrumentation and accuracy, or use established methodology to sample stream invertebrates (Arthurs 2019, Scott et al. 2012). In addition to considering these three fundamental learning domains, UFEs may promote student outcomes that span domains and enter the social realm, such as developing communication skills (Bell and Anscombe 2013), building friendships and collaborations (Stokes and Boyle 2009, Jolley et al. 2019), and/or developing a sense of belonging in a discipline (Kortz et al. 2020, Malm et al. 2020, O'Brien et al. 2020). Lastly, students participating in UFEs could result in broader, societal level outcomes, such as students pursuing conservation efforts, contributing to citizen science projects, increasing awareness for social justice issues, or supporting for sustainability efforts (Grimberg et al. 2008, Bell and Anscombe 2013, Ginwright and Cammarota 2015).

In **Table 1**, we present a list of common intended student outcomes from UFEs. The list of outcomes was propagated by UFE practitioners, first identified from a UFERN landscape study (O'Connell et al. 2020) and by participants at the 2018 UFERN meeting. O'Connell et al. (2020) surveyed practitioners on expected student outcomes from their UFEs. We then refined the list of outcomes by removing outcomes that were redundant, not measurable, or linked to very specific contexts (not field universal), and then grouped them by what we call 'primary aim'. The primary aim category is an umbrella category by which to group similar intended outcomes. **Table 1** illustrates a diversity of possible and likely outcomes from UFEs ranging across domains, but not every conceivable outcome is accounted for, and we encourage practitioners to consider outcomes that they do not see on this table if they are in alignment with their UFE. Interestingly, in O'Connell et al.'s (2020) survey of intended student outcomes in extended UFEs, the majority of respondents chose outcomes in the cognitive and/or psychomotor domains. Thus, students gaining content knowledge and skills is a prominent goal for practitioners of UFEs, but content can also be learned in many contexts. We and others propose that the distinctive impact of participation in a UFE may actually be more in the affective domain (Van Der Hoeven Kraft et al. 2011, Kortz et al. 2020). Thus, we encourage practitioners to consider focusing less on content level outcomes and more on the full spectrum of possible outcomes.

II. Consider the Context of the UFE

UFEs can be highly variable in format (Lonergan and Andresen 1988, Whitmeyer et al. 2009b, O'Connell et al. 2020). For example, some are strictly disciplinary (Jolley et al. 2018b), others interdisciplinary (Alagona and Simon 2010); they might occur locally (Peacock et al. 2018), in short duration (Hughes 2016), over an entire course (Thomas and Roberts 2009), or as a summer research experience held at a residential field station (Hodder 2009, Wilson et al. 2018). O'Connell et al., (2021) comprehensively describes and organizes the evidence for how student factors such as student identity, prior knowledge, and prior experience and design factors such as setting and social interaction influence learning in the variety of UFE formats (O'Connell et al., in press). In this paper, we urge practitioners to consider student factors (e.g. prior knowledge, skills and experiences, motivation and expectations, social identity, and personal needs) and design factors (e.g. setting, timing, instructional models and activities) when determining an appropriate assessment approach. These contextual factors should inform assessment decisions as well as data interpretation, and how to use the data to make decisions about next steps in assessment or evaluation. The intention is for practitioners to use the guide (Fig. 1) to inform iterative change and improvement and reflective practice, not as static scaffolding.

Student Factors

As with any learning environment, it is critical for instructors and staff to have a good idea of who the participating students are, and preempt what information may be pertinent to their experiences as practitioners plan to understand the outcomes of a UFE (Pender et al. 2010, Fakayode et al. 2014, Ireland et al. 2018, Stokes et al. 2019). In this way, student factors may influence the selection of appropriate assessment approaches and tools. There are a number of factors that can be considered when designing and understanding the outcomes of assessment; here we provide numerous examples for contemplation.

For example, a factor to consider is prior student knowledge and skills. Imagine two UFEs: in the first UFE, students are upper-division physiology majors studying endemic amphibians' responses to changes in stream water quality; the second UFE is designed for non-science majors to broadly survey the biodiversity of local flora and fauna. If a practitioner decides they want to identify if/how students' attitudes change regarding the local environment as a result of the UFEs they might select a survey designed to collect data on environmental attitudes (e.g. Table 1, Primary Aim: Connection to Place; Assessment Tool: Environmental Attitudes Inventory (EAI), Milfont and Duckitt 2010). The physiology students from the first example may begin the UFE with largely positive environmental attitudes already. Thus, administering a survey at the beginning and end of the UFE (pre-post) to measure this construct may not reveal any gains. Yet, in the second UFE example, the students are introductory, non-science majors, and they may demonstrate significant, quantifiable gains in environmental attitudes. Therefore, in the physiology student example, this specific outcome was not detectable due to a measurement limitation called the ceiling effect. This effect can occur when a large proportion of subjects begin a study with very high scores on the measured variable(s), such that participation in an educational experience yields no significant gains among these learners (Austin and Brunner 2003, Judson 2012). In this case, instead of the survey, the practitioner might learn more by crafting an essay assignment that probes the physiology students' environmental values. This option would demonstrate consideration of the student population in the assessment strategy.

Other factors to consider might include student motivation and expectations. An assessment of students in a pair of geoscience UFEs in New Zealand showed that study abroad students were more intrinsically motivated, pro-environmental, and had a stronger sense of place than local students in a similar field experience, although they were held in the same place (Jolley et al. 2018a). This assessment highlighted the need to adapt the design of the field experience to be more applied, environmentally focused, and place-based, rather than simply applying the same curricula unchanged to a different student population (Jolley et al. 2018a). Here, future assessments could be targeted towards investigating whether the revised UFE design for study abroad students effectively captured their motivation and interest. And/or, a deeper qualitative investigation could be conducted to characterize their field experiences in relation to the environmental and place-based content.

Prior experiences and identity are also critical to consider (Scott et al. 2019, Morales et al. 2020). Have the students experienced fieldwork already? Practitioners might want to know what proportion of the students are first-generation college students, or if students have prior conceptions of fieldwork. Such knowledge could guide an assessment approach aimed at understanding how first-generation students experience the UFE compared to continuing generation students; or in the latter case, if students hold accurate inaccurate conceptions (or any conception at all) about fieldwork.

Also important is awareness of personal needs such as safety and well-being, especially for students of often marginalized identities such as BIPOC (Black, Indigenous, and People of Color) students and LGBTQ+ students (John and Khan 2018, Anadu et al. 2020, Giles et al., 2020; Marín-Spiotta et al. 2020, Demery and Pipkin 2021). These considerations can influence the implementation of an assessment strategy, as participants will experience different levels of comfort and risk based on the questions being asked. Students may be less comfortable sharing if they already have concerns about safety in the field environment and culture of UFEs. Even on an anonymous survey, students may be worried about being personally identifiable if they are one of few students of a particular identity or combination of identities. Ensure that students are provided full and complete information about what will be done with their data, have the opportunity to ask questions, and are free from coercion. In some cases, this may mean having someone who is not the course instructor conduct the assessment. Although questions like these would be addressed if the study requires approval through an IRB or similar, we encourage their consideration regardless as they have a bearing on student comfort and perceptions of safety.

Programmatic processes such as recruitment efforts or selection criteria can also influence student factors (e.g., Zavaleta et al. 2020, O'Connell et 2021). Are all students enrolled in a class participating in the UFE (as in a CURE), do they self-select, or are they chosen to participate based on certain criteria? It is important to keep in mind that any outcomes from a UFE are only representative of the students who actually participated, and thus not broadly representative of any student who might participate. In summary, when applying the assessment strategy presented in this paper, one must consider: Are the UFE outcomes reasonable to achieve and measure given the specific student population? Student factors must be considered in UFE design and will likely moderate or even become the subject of assessment efforts.

In the vignettes, we identify various factors that may inform program design/UFEs and provide diverse examples in which the assessment approaches are aligned with the student population. For example, some programs specifically engage students with a background or interest in STEM (e.g., **Fig.2A, 2B**), others are open to all majors (e.g. **Fig. 2C**).

Setting and Timing

Fundamental to the definition of UFEs is that they are immersive, communal, and somewhat unstructured (even if conducted remotely) (Posselt et al. 2020, p. 56-57). This distinctive learning environment should be considered when picking an assessment approach and interpreting assessment data. If a practitioner wanted to evaluate how a UFE impacts student knowledge of a particular concept, then a two-week, on-campus UFE focused on urban greenspaces may yield less deep learning about forest ecology than a semester-long field course held in a live-in forest field station. Thus, a summative assessment on forest ecology concepts should be reflective of the amount of time and depth the students have had to amass relevant cognitive gains.

Previous work indicates that instructors and students place high value on UFEs where participants live and work together in the field (Jolley et al. 2019). However, cohabitation and isolation may also present challenges in the way of mental health stressors (John and Khan 2018) and unfamiliar and overstimulating environments (Kingsbury et al. 2020). In an almost opposite, yet timely and relevant example, Barton (2020) describes how remote UFEs need to reduce or change expected learning outcomes specific to being "in the field" to outcomes more relevant. Considering how the UFE setting might impact student learning should be factored into determining intended student outcomes, and subsequently how to test if those outcomes are being met. **Fig. 2** illustrates how factors such as residential/non-residential settings, length of the UFE, and accessibility of the setting can inform assessment strategies.

Contextual Factors Can Intersect

The student experience (and thus the student outcomes) are influenced by the intersection of setting and timing factors, making interpretation of the results complex. For example, perhaps a student is a primary caregiver for someone at home and is distracted by irregular or absent cellular service, therefore are unable to establish a connection to place due to distraction and worry. Some students may identify that eating as a community helps them to establish a sense of belonging among peers and instructors, whereas, eating in a group setting may cause a student with a complex relationship with food to experience extreme discomfort. These examples are provided highlight how residential or community settings may have contradictory impacts on different students in the same UFE, thus it may not always be appropriate or meaningful to solely look at assessment findings on an average or "whole-class" scale.

Instructional Model and Activities

As with any learning experience, working backwards from the specific learning outcomes will help instructors to ascertain if the curriculum is in alignment with those goals, or if there are activities that are not aligned or extraneous. If intended student outcomes are to increase skills with research practices (e.g. **Fig. 2A**), then the actual activities should support this outcome. In this vignette, students are supported to develop a research project, aligning the instructional model and activities to the outcome. Similarly, an intended outcome of the *Humanities Course at a Field Station* vignette (**Fig. 2C**) was to develop stronger connections to place in Northern Michigan, and the course curriculum included activities focused on exposure to place, and fostering a sense of place. In the *Urban Field CURE* vignette (**Fig. 2B**), an intended outcome was for students to engage with relevant stakeholders, and activities included gaining feedback on student-developed experimental design from the researcher's whose work the urban field CURE expanded. There are multiple options for designing curriculum or activities that will allow practitioners to gauge the participant experience, thus acting as a form of formative assessment. For example, designing a written reflection activity that probes the student experience or their learning in that particular environment, or collecting student artifacts from the field experience can yield information regarding how a student experiences the UFE, and can in turn inform UFE stakeholders.

Accessibility and Inclusion

As illustrated previously, basic characteristics of the location and pedagogy of the UFE can have an impact on the physical, cognitive, and/or or emotional accessibility of the learning environment for various students. In efforts to include as many students as possible, it is important to consider factors such as physical space (e.g., restroom availability, non-gendered housing, housing for students with physical, emotional, or psychological concerns), quality of internet connection (if remote), sleeping arrangements, skills needed to participate (e.g., training in swimming), or other health concerns (e.g., allergies). Additionally, social isolation/inclusion can be especially prevalent in UFEs for students who don't share the same identities with previous participants and/or are from underrepresented groups (Atchison et al. 2019, Morales et al. 2020). **One of the vignettes (Fig. 2D)** is specifically tied to accessibility*,* and demonstrates the importance of directly working with students and faculty with disabilities on a field trip in order to address the intended outcomes of the UFE.

III. Assessment Approach

Key to choosing an assessment approach is first asking: What is the motivation for collecting the data? As discussed earlier, there are a number of reasons and ways one might assess a UFE including identifying if students are meeting specific learning goals, to collect publishable data on students' sustained interest in a topic, or to identify if the UFE is meeting programmatic goals in order to report back to a funding agency or university. Regardless of stakeholders' motivations, using backward design to clarify and align program goals, activities and assessments will allow for a solid platform for improvement and evaluation.

We recommend that practitioners consider both formative and summative assessments. A formative assessment might be a UFE student completing a written reflection or keeping a "reflective diary" (Maskall and Stokes, 2008, Scott et al. 2019) regarding an aspect of their learning experience. This strategy would provide students a chance to reflect on their learning process and their changing experience and competencies in their own words. Further, such a formative assessment would allow instructors/stakeholders to better understand how programming, or more specifically a particular aspect of programming may impact student perceptions and possibly how to adjust the learning experience. A summative assessment strategy could be employed if practitioners wanted to know if students have gained a greater appreciation for the natural world as a result of a UFE, which could be measured for example by conducting a pre/post survey designed to measure this specific construct (e.g. Table 1. Primary Aim: Connection to Place, Assessment Tool: Place Attachment Inventory (PAI), Williams and Vaske 2003). Fig. 1 is meant to be useful in planning assessment strategies but could also serve as a helpful communication tool when engaging with funders and stakeholders.

It may also be appropriate to hire an external evaluator. An advantage of external evaluation is that it presumably provides an unbiased view of the program, as the evaluator will assess the impacts of programming on participants and report findings in an objective manner. From the evaluator's perspective, is the program meeting its intended goals? For whom does the UFE appear to be "working", and are there certain student groups that are not being impacted in the way designers of the experience had intended? An external evaluator will often work with the team to identify goals, and then conduct a holistic programmatic evaluation, including all stakeholders. The caveat regarding external evaluation is cost. If grant-funded, external evaluation may be encouraged or even required; if not grant-funded, finding funding would be necessary in order to hire the evaluator or evaluation team.

Data Collection and Analysis

Deciding what type of data to collect will require having a reasonable idea of the program's goals and anticipated outcomes, as well as an awareness of the time it will take to collect and then analyze the type of data collected. Practitioners may consider using quantitative measures such as surveys, or qualitative methods such as interviews or open-ended questions. A mixed methods approach can employ both qualitative and quantitative methodology, allowing for a more nuanced understanding (Creswell and Clark 2007). Identifying if the intention is to publish the data (requiring IRB review), or to use it internally to gain a better understanding of an aspect of programming should play a key role in determining the approach and the 'rigor' with which one collects and interprets the data.

Using best practices in research will help aid in avoiding conflicts of interest, and better ensure that valid and reliable data is collected (Ryan et al. 2009). If, for example, a program recruits students for interviews after they participate in a UFE, someone outside of the UFE leadership or instructional team should be the interviewer. This practice would help to minimize the power differential between participant and researcher, thereby ensuring that UFE interview participants feel that they can be honest about their experiences, and not worry about pleasing or offending those involved in the program (Kvale and Brinkman 2009). Further, the interview questions should be vetted by others (similar to target audience) before the interviews begin to ensure that the questions are interpreted by the participants as intended and appropriate for the specific student population.

Using appropriate methodology in planning data collection and conducting analyses, will allow for apt interpretation of the results (Clift and Brady 2005). As illustrated in the vignettes (**Fig. 2D**), deeply understanding the lived experiences of participants may call for knowledge of qualitative methodology. One may not want to conduct numerous interviews with students and staff without the resources to hire researchers, or ample time to analyze the data. Analyzing rich qualitative data typically involves iterative "coding" by multiple trained researchers who develop and revise codebooks and then apply those codes to the transcribed text, regularly checking for coding reliability among researchers (Saldaña 2011, Belotto 2018, O'Connor and Joffe 2020). Coding processes can vary, sometimes guided by a theoretical framework, *a priori* ideas, and/or they may allow for inductive, deductive or a combination of coding approaches (see Saldaña 2015 for a comprehensive manual on coding).

Similar to qualitative data, quantitative data collection and analysis requires planning and expertise. Researchers will want to ensure that the research aims are well-aligned with the data collection methods or tools, and in turn, allow for appropriate interpretation the data. Comparing pre-post survey responses would be one seemingly straightforward way to measure change over time in participant learning (e.g., **Fig. 2C**). Yet, we do caution against simply pulling a tool from Table 1 or elsewhere and simply assuming that by using it, it 'worked'. We recommend collaborating with experts who are familiar quantitative methods. Using a survey tool may yield quickly quantifiable results, but if the survey has not undergone vetting with individuals similar to the population of study, or it has not previously shown to collect valid data in very similar populations, one cannot assume that data collected is valid or reliable (Fink and Litwin 1995, Barbera and VandenPlas 2011). Just as we do not use micropipettes to measure large volumes of lake water, we would not use a tool developed to measure academic motivation in suburban elementary school students to measure motivation of college students participating in a residential UFE and expect to trust the survey results outright. If a tool seems appropriate for a given UFE and the student population, we encourage first testing the tool in that population and work to interpret the results using best practices (for a comprehensive resource on these practices, see American Educational Research Association (AERA) 2014). As described previously, **Table 1** consists of several assessment tools which are potentially relevant for measuring UFE outcomes. We only included tools that have been peer-reviewed and published in the table. We strongly recommend reviewing the associated peer-reviewed paper before using a tool, as well as looking in the literature to see if others have used the tool and published their findings.

It is also possible that one would want to measure an outcome for which a tool has not yet been developed. In this case, working on an attuned assessment strategy based on iterative adaptations and using lessons learned may be appropriate (Adams and Wieman 2011). There are many steps involved with designing and testing a new assessment tool that is capable of collecting valid and reliable data. Therefore, if stakeholders deem it necessary to create a new tool to measure a particular outcome, or develop or modify theory based on an UFE, we recommend working with psychometricians or education researchers.

IV. What are the Next Steps?

We encourage that the process of evaluation and assessment is a reflective, cyclical, iterative process of improvement as it relates to UFE design and implementation. There are inevitably going to be aspects of any learning experience that could be improved, and this guide to assessment (Fig. 1) can help practitioners visualize alignment between intended outcomes, programming, assessment and evaluation; and how each informs the other. The next steps for many UFEs might be to first report to stakeholders (funders, the institution, etc.) on the outcomes of the UFE. Or, if the goal of the assessment effort was to conduct novel research, then the next steps might be to analyze, write up and submit the results of the study for peer review, thereby contributing to the growing literature of empirical outcomes from UFEs. For example, one vignette (**Fig. 2B)** describes how the assessment strategy will provide pilot data for ongoing publishable projects. Other vignettes **(Fig. 2A&C)** illustrate how results from assessment efforts can be leveraged to apply for or validate grant funding. These types of data may be paramount to sustained funding, data-driven advocacy efforts, and/or applying for future funding for continued programming.

An important part of the presented strategy is that it might be used to engage stakeholders in a discussion about what additional questions might be appropriate to ask or what improvements need to be considered. Is there alignment between activities and learning goals? Is the current evaluation strategy accurately measuring what stakeholders expect the students to gain from the UFE? Is the programing intentionally inclusive of the participants' diverse perspectives and experiences, or could adaptations be made to better serve the UFE population? For example, to address financial and relocation barriers identified through the program evaluation for one field based REU, the REU leaders introduced new policies for students to be paid at the start of their experience and identified field research projects that were located in student communities, and in another case, accommodations were made for the student's family to join them as part of the residential field experience (Ward et al. 2018). This is just one example of how assessment data can be used to inform the design of future UFEs and highlights how the assessment process can be both informative and iterative.

EXPANDED VIGNETTES

Here we provide detailed narratives that more fully illustrate two of the vignettes introduced in Fig. 2 (Fig. 2A, C). The expanded vignettes are intended to transform the collective ideas presented here and summarized in Fig. 1 into concrete examples, serving as an example to guide assessment of diverse UFEs.

A

Summer Research Experience for Undergraduate Students

Outcomes: (1) Develop research skills; (2) Build knowledge of discipline-specific concepts and content; (3) Strengthen proficiency with discipline-specific methods and procedures.

Context: Residential, 10 weeks. Undergraduate students, typically STEM majors, with interest in ecological and/or environmental fields and who have completed pre-requisite courses. Students are assigned to work in small groups on a long-standing project at a biology field station. In addition, they are required to develop a small-scale, independent-study project. At the end of the summer, students are required to deliver an oral presentation on their assigned project and a poster presentation on their independent-study project.

Approach: Formative and summative evaluation for program refinement, internal requirements and external support. Analysis of student research products (oral and poster presentations).

Next Steps: Pre-experience participant assessments, addressing all learning outcomes. Add additional student outcomes and seek appropriate assessment tools. Evaluate program outcomes for improvements for future offerings, internal institutional requirements and support from external funders.

B

Urban Field CURE: Course-Based Undergraduate Research Experience

Outcomes: (1) Engage with stakeholders in a locally relevant, urban CURE using scientific practices, iteration, collaboration, and broader relevance/discovery; (2) Identify local and non-native plants and analyze their relationship to air pollution and urban landscapes; (3) Refine and articulate career goals.

Context: Non-residential, 3-10 weeks. Upper-division biology and environmental science undergraduates. Occurs in urban field sites and on-campus. Students work as a group to consult with either research or parks/forest service stakeholders in the design and execution of a novel research project. Students present their findings to the stakeholders.

Approach: For instructor knowledge and internal evaluation of efficacy of field CUREs on self-reported skills, research abilities, presentation skills and future plans. Data include: student coursework, stakeholder written evaluations, presentation evaluations and an open-ended survey on career goals and future plans.

Next Steps: Based on student products and their contribution to stakeholder research, assess if the research can continue independently, or could be repeated for more data and eventually publication. Compile student reported outcomes to present to departments for future funding and support for more field CUREs.

C

Humanities Course at a Field Station

Outcomes: (1) Develop stronger connections to place in northern Michigan; (2) Strengthen ability to communicate with/about scientific work; (3) Enhance value for the interdisciplinary nature of science.

Context: Six-week residential field camp in remote/unpopulated setting. Five other classes running at the same time. Program open to students in all majors and room and board was free first year of program. Curriculum was place-based, focused on learning from and not just about. Students involved in this program take four courses; each course is taught by a different instructor.

Approach: Mixed methods assessment. Used pre-post module survey with previously validated instruments (Place Attachment Inventory) and focus groups to investigate perceived scientific communication and value for interdisciplinary nature of science.

Next Steps: Conduct interviews to explore how connection to each other (sense of belonging in class) impacts student learning outcomes. Share assessment findings with instructors and TAs to determine program modifications. Provide evaluation of program to funding agency and collect evidence for future institutional or external support of program.

D

An Accessible Field Experience for Students and Instructors

Outcomes: (1) Enable all participants to explore the regional geology of the field area; (2) Engage the abilities of all participants in the field; (3) Collaboratively identify effective practices for accessibility and inclusion in field-based teaching and learning.

Context: Non-residental, one day field trip in a workshop format. An equal number of student and instructor participants, with and without disabilities. Multiple field sites visited, with a worksheet to be completed on the content, alongside broader discussions of accessibility and inclusion in the field.

Approach: Extensive qualitative research on the lived experiences of participants, through interviews and observations. Expertise in qualitative data required for interpretation.

Next Steps: Conduct similar research in other field trips and contexts (e.g., different locations, multi-day). Share recommendations with community to improve the access of all field experiences.

Figure 2. Vignettes of Undergraduate Field Experiences (UFEs). These vignettes (**A-D**) represent actual examples of UFEs and illustrate how to apply the components of **Fig. 1** (Strategy for Assessment of Undergraduate Field Experiences (UFEs)) to assess each UFE. Figure 2D was based on (Gilley et al. 2015, Feig et al. 2019, Stokes et al. 2019).

Vignette A – Summer Research Experience for Undergraduate Students (Fig. 2A)

The field site and course: The Thomas More University (TMU) Biology Field Station was founded in 1967 and offers research, courses, and field experience programs for undergraduate students and outreach programs for K-12 students and the general public. The TMU Biology Field Station is located 20 miles from the main campus in a more remote/unpopulated setting, along the banks of the Ohio River. Each summer, undergraduate students from around the country are selected to participate in a 10-week summer research internship where they are assigned to one of three long-standing research projects and develop an independent-study side project on which to develop and work throughout the ten weeks.

Development of student outcomes: During the preceding academic year, TMU Biology Field Station staff, including the field station director, discussed outcomes that they wanted to achieve with these internships. These outcomes were informed by discussions with the faculty from the Department of Biological Sciences at TMU and with collaborating researchers at the Environmental Protection Agency (EPA) Office of Research and Development and the US Fish and Wildlife Services (USFW). The primary, intended student outcomes included (1) Increased understanding of and proficiency with research practices and processes; (2) Increased understanding of discipline-specific concepts and content; and (3) Stronger skills in discipline-specific methods and procedures. Secondary student outcomes included (1) Expanded professional networks; (2) Greater sense of belonging in the scientific community; (3) More refined career goals; and (4) Stronger professional skills.

Course and station context: To qualify, students must have completed one year of general biology and/or one year of general chemistry while maintaining a 3.0 minimum GPA. The qualifications to apply are kept at a minimum, by design, to ensure that first-year students are eligible to apply. No prior research experience was required. The application process was open in December; applications were due in early February; and selections were made in early March for the subsequent summer. Phone or face-to-face interviews were conducted with each finalist as part of the application process. All interns were required to live on site. A stipend and free housing were provided.

During the internship, students were assigned to one of three long-term projects at the TMU Biology Field Station and conducted this research as part of a small group of students and one faculty mentor. In addition, students were required to conduct a small-scale independent-study project of their own choosing, in collaboration with a faculty mentor. For the independent-study project, students were required to conduct a literature search, write a proposal and carry out the project within the course of their summer internship. At the conclusion of the summer, students made on oral presentation on their group work and a poster presentation on their independent project.

In addition, student interns were required to attend a summer seminar series during which professionals presented their research and spent a day observing the students in action. Lastly, students participated in field trips and tours to labs at the EPA, USFW, and local governmental agencies and served as mentors for a weeklong STEM camp for high school students.

The TMU Biology Field Station is a residential field station, where students live together in houses. In addition to the residential structures, there are three labs, four classrooms and a STEM Outreach Center. Students, staff and faculty eat meals together and socialize together in both formal and informal activities throughout the summer.

Data collection: In order to assess change (increases in perceived ability or value), the field station director used a pre/post survey to identify student perceptions before they began the internship and after they ended the internship. The survey included measures about research practices and processes, discipline-specific concepts and content, and discipline-specific methods and procedures. The survey also included measures about career goals and professional skills. The field station director also conducted mid-summer and exit interviews with each student intern to explore perceptions about their knowledge and skills gained through the program. While this assessment was created for an institutional annual report, the Director also used these data for support of additional external funding in grant applications and also compared the findings to previous years' surveys.

Next steps: Findings from the survey responses and interviews indicated that students in the internship program gained knowledge and skills in research practices and in discipline-specific content, methods and procedures. Further, students indicated more refined career goals and professional skills, namely oral and written skills. Students in the internship perceived increased confidence in their ability to communicate about science and an increased scientific network.

Future assessment work will consist of additional surveys and interviews with students a year later to explore how the internship experience impacted their academic work in the subsequent school year and career development. Lastly, attempts are being made to contact student interns from previous years to determine their specific career path and status.

Vignette C – Humanities Course at a Field Station (Fig. 2C)

The field site and course: University of Michigan Biological Station (UMBS), which was founded in 1909, houses research, courses, and field experience programs for students. UMBS is located 250 miles from central campus in a remote setting. The *Humanities Course at a Field Station* was a newly designed course which was part of a larger effort to bring students from other disciplines to UMBS.

Development of student outcomes: During the humanities course development, UMBS staff, including the program manager and program evaluation coordinator, discussed outcomes that they wanted to explore with this particular class to include in their annual program assessment. These outcomes were informed by discussions with the faculty as well as through reviewing syllabi. The intended student outcomes included (1) Develop stronger connections to place in northern Michigan; (2) Increased ability to communicate about scientific work; (3) Increased value for the interdisciplinary nature of science.

Course and station context: The humanities course was open to all undergraduate students across majors, room and board was free for the first year of the program for students, scholarship assistance was available, and transportation was provided. The course ran for six weeks during the UM spring term, which allowed students opportunities to work or take other courses during the rest of the summer. The course was a place-based course, where the focus was on learning from the place and not just about the place. Students involved in this course took four short courses and received 8 credit hours across three departments (English, Anthropology, and American Culture); each course was taught by a different instructor.

UMBS is a residential field station, where students live together in cabins and faculty also live on-site. Students and faculty eat meals together in the dining hall. Five other undergraduate courses ran at the same time as the humanities course. These additional five courses came from more traditional biophysical disciplines such as general ecology and biology of birds. While students in the humanities course generally spent time with their classmates and faculty in their individual course, there were opportunities (both structured and unstructured) for students to communicate, work with, and form connections with students, researchers, and faculty in other courses.

Data collection: In order to assess change (increases in perceived ability or value), the program evaluation coordinator used a pre/post survey to identify student perceptions before they began the course and after they ended the course. The survey included measures about sense of place, sense of connection to larger-scale problems or issues, and ability to communicate with scientists about scientific work. The program evaluation coordinator also conducted a focus group with students in the course to explore perceptions about their value of the interdisciplinary nature of science, ability to communicate, and connections to place in more detail. Interviews with the instructor and a focus group with the TA for the course also provided insight into change in student perceptions about these topics and *how* these changes developed in their time taking this course at UMBS.

While this assessment was created to share for an annual report, the program evaluation coordinator was interested in sharing this information with the larger field education community, and so all of the assessment of this course (and all courses at UMBS) had IRB approval. In addition, the program evaluation coordinator selected published measures to include on pre/post surveys that had been tested in college populations. The program evaluation coordinator intentionally conducted focus groups because students had no interaction with her until this meeting and she was not associated with their grades or evaluation for their course.

Next steps: Findings from the first year of survey responses and focus groups indicated that students in the course formed extremely close-knit bonds. Future assessment work will consist of interviews with students, faculty, and TA to explore how connections to others (sense of belonging in the class) impacts learning and understanding of different course topics.

In addition, findings from surveys and focus groups indicated that students in the course perceived increases in the value of the interdisciplinary nature of science and increased confidence in their ability to communicate about science. Findings from faculty interviews supported student responses and also indicated that faculty had a strong interest in doing more intentional collaboration with biophysical courses in the future. After discussing all of the assessment data, UMBS staff decided to expand their assessment for the next year. Specifically, they wanted to know if students from bio-physical courses who interacted with students in the humanities course also experienced increases in perceived value of the interdisciplinary nature of science and ability to communicate about science. The program evaluation coordinator intends to add additional assessment approaches to examine interactions between this course and other courses at the station. This may include observations of structured and unstructured activities with the humanities and bio-physical courses as well as adding survey questions and/or focus group questions for all students who are taking courses at UMBS. Thus, the results of the assessment of the humanities course not only addressed whether the student outcomes were achieved in the humanities course, but also highlighted changes in the program that would happen in future iterations, and informed additional assessment of all UMBS courses in the next year.

Conclusions

We encourage using contextual information about a UFE to iteratively inform assessment strategies and in turn, improve the value and inclusivity of the UFE for the full spectrum of participants and stakeholders. We encourage practitioners to use the supports provided here to conduct applied research aiming to understand how various characteristics of UFEs impact various student populations, essentially to "identify what works for whom and under what conditions." (Dolan 2015, National Academies of Sciences and Medicine (NASEM), 2017) p. 175). In general, we have little empirical evidence about the linkage of program characteristics to learning outcomes in UFEs. O'Connell et al., (2021) presents an evidence-based model that hypothesizes how *student context factors* and *program design factors* (or program characteristics) impact student outcomes in UFEs. Through a thoughtful assessment approach along with consideration of student context factors, practitioners may begin to unravel which design factors of their UFE are specifically leading to which student outcomes for which students. Future work could model which design factors lead to specific outcomes, as demonstrated by work to better understand how CURE elements influence student outcomes (Corwin et al. 2015).

We believe that the process of informed assessment and reflection will improve the accessibility and inclusivity of UFEs. Morales et al. (2020, p. 7) call for continuing a "*conversation about creating student-centered field experiences that represent positive and formative experiences for all participants while removing real or imagined barriers to any student participating in field research*." Explicit attention to diversity, equity, access, and inclusion regarding who gets to participate in UFEs and the learning that results from the experiences, are key conversations with important implications (Nairn 1999, Carabajal et al. 2017, Stokes et al. 2019, Giles et al. 2020, Morales 2020 et al., Zavaleta et al. 2020, Demery and Pipkin 2021). As illustrated in Fig.2D for example, authentically considering what it means to be accessible and inclusive is an important question, and we suggest that practitioners begin to systematically evaluate who is served by their UFE and who is not served and why, thus deeply investigating how the UFE may become more inclusive for diverse individuals. It will be necessary to work across disciplines to learn what is needed to support and advocate for accessible and inclusive UFEs such that as many students as possible can participate and have a positive experience.

The recent COVID-19 pandemic has brought to the forefront vital questions about the role of virtual field experiences (Arthurs 2021, Swing et al. 2021), as well as assessment practices that are in alignment with these. We suggest that this is one area where developing novel assessment tools are needed to effectively measure impact and to ask such questions as: What are the characteristics defining a virtual UFE? As it relates to outcomes, what can we learn about the impacts of in-person experiences vs. remote on a student's affect such as their sense of belonging?

Here we meet a call from the community to aid practitioners and stakeholders in using best practices to assess, evaluate, and/or research the spectrum of UFEs. UFEs are widespread and diverse, yet unique and complex. As we consider more deeply the outcomes that are specific to UFEs, we urge practitioners to move towards evidence-based advocacy and improvement for the continued support of UFEs.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation under RCN-UBE grant #1730756 (Awarded to K.O.). We would like to thank participants of a UFERN panel and Erin Dolan for their helpful feedback on previous drafts of this paper.

AUTHOR CONTRIBUTIONS

All authors conceived of, wrote, and revised the paper.

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA ACCESSIBLITY STATEMENT

N/A

REFERENCES

Adams, W. K., and C. E. Wieman. 2011. Development and Validation of Instruments to Measure Learning of Expert-Like Thinking. International Journal of Science Education **33**:1289-1312.

Alagona, P. S., and G. L. Simon. 2010. The role of field study in humanistic and interdisciplinary environmental education. Journal of experiential education **32**:191-206.

American Educational Research Association (AERA), A. P. A., and National Council on Measurement in Education, eds. 2014. Standards for educational and psychological testing. American Educational Research Association.

Anadu, J., H. Ali, and C. Jackson. 2020. Ten steps to protect BIPOC scholars in the field. Eos, 101, https://doi. org/10.1029.

Anderson, L. W., D. R. Krathwohl, and B. S. Bloom. 2001. A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Allyn & Bacon.

Arthurs, L. A. 2019. Undergraduate geoscience education research: Evolution of an emerging field of discipline-based education research. Journal of Research in Science Teaching **56**:118-140.

Arthurs, L. A. 2021. Bringing the Field to Students during COVID-19 and Beyond. GSA Today.

Atchison, C. L., A. M. Marshall, and T. D. Collins. 2019. A multiple case study of inclusive learning communities enabling active participation in geoscience field courses for students with physical disabilities. Journal of Geoscience Education **67**:472-486.

Austin, P. C., and L. J. Brunner. 2003. Type I error inflation in the presence of a ceiling effect. The American Statistician **57**:97-104.

Barbera, J., and J. VandenPlas. 2011. All Assessment Materials Are Not Created Equal: The Myths about Instrument Development, Validity, and Reliability. American Chemical Society.

Barker, S., D. Slingsby, and S. Tilling. 2002. Teaching biology outside the classroom. Is it heading for extinction:14-19.

Barton, D. C. 2020. Impacts of the COVID-19 pandemic on field instruction and remote teaching alternatives: Results from a survey of instructors. *Ecology and evolution*, *10*(22), 12499-12507.

Bauerle, T. L., and T. D. Park. 2012. Experiential learning enhances student knowledge retention in the plant sciences. HortTechnology **22**:715-718.

Bell, K., and A. Anscombe. 2013. International field experience in social work: Outcomes of a short-term study abroad programme to India. Social Work Education **32**:1032-1047.

Belotto, M. J. 2018. Data analysis methods for qualitative research: Managing the challenges of coding, interrater reliability, and thematic analysis. Qualitative Report **23**.

Billick, I., and M. V. Price. 2010. The ecology of place: contributions of place-based research to ecological understanding. University of Chicago Press.

Bloom, B. S., and D. R. Krathwohl. 1966. Taxonomy of educational objectives. Handbook I: Cognitive domain.

Boyle, A., S. Maguire, A. Martin, C. Milsom, R. Nash, S. Rawlinson, A. Turner, S. Wurthmann, and S. Conchie. 2007. Fieldwork is good: The student perception and the affective domain. Journal of Geography in Higher Education **31**:299-317.

Brewer, C. A., and D. Smith. 2011. Vision and change in undergraduate biology education: a call to action. American Association for the Advancement of Science, Washington, DC.

Bruening, M., S. Brennhofer, I. Van Woerden, M. Todd, and M. Laska. 2016. Factors related to the high rates of food insecurity among diverse, urban college freshmen. Journal of the Academy of Nutrition and Dietetics **116**:1450-1457.

Cajete, G. 2000. Native science: Natural laws of interdependence. Clear Light Pub.

Cameron, A. K. 2007. Cultural and religious barriers to learning in basic astronomy: A South African study.

Carabajal, I. G., A. M. Marshall, and C. L. Atchison. 2017. A synthesis of instructional strategies in geoscience education literature that address barriers to inclusion for students with disabilities. Journal of Geoscience Education **65**:531-541.

Cervato, C., and R. Frodeman. 2012. The significance of geologic time: Cultural, educational, and economic frameworks. Geological Society of America Special Papers **486**:19.

Clift, R. T., and P. Brady. 2005. Research on methods courses and field experiences. Studying teacher education: The report of the AERA panel on research and teacher education **309424**.

Corwin, L. A., M. J. Graham, and E. L. Dolan. 2015. Modeling course-based undergraduate research experiences: An agenda for future research and evaluation. CBE—Life Sciences Education **14**:es1.

Creswell, J. W., and V. L. P. Clark. 2007. Designing and conducting mixed methods research.

Creswell, J. W. 2013. Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

Cutter, E. G. 1993. Fieldwork: an essential component of biological training.

Dayton, P. K., and E. Sala. 2001. Natural history: the sense of wonder, creativity and progress in ecology. Scientia Marina **65**:199-206.

Demery, A.-J. C., and M. A. Pipkin. 2021. Safe fieldwork strategies for at-risk individuals, their supervisors and institutions. Nature Ecology & Evolution **5**:5-9.

Dolan, E. L. 2015. Biology Education Research 2.0. CBE Life Sciences Education **14**.

Dressen, D. F. 2002. Accounting for fieldwork in three areas of modern geology: A situated analysis of textual silence and salience. Citeseer.

Easton, E., and A. Gilburn. 2012. The field course effect: Gains in cognitive learning in undergraduate biology students following a field course. Journal of Biological Education **46**:29-35.

Fakayode, S. O., M. Yakubu, O. M. Adeyeye, D. A. Pollard, and A. K. Mohammed. 2014. Promoting undergraduate STEM education at a historically black college and university through research experience. Journal of Chemical Education **91**:662-665.

Feig, A. D., C. Atchison, A. Stokes, and B. Gilley. 2019. Achieving Inclusive Field-based Education: Results and Recommendations from an Accessible Geoscience Field Trip. Journal of the Scholarship of Teaching and Learning **19**.

Fink, A., and M. S. Litwin. 1995. How to measure survey reliability and validity. Sage.

Fleischner, T. L., R. E. Espinoza, G. A. Gerrish, H. W. Greene, R. W. Kimmerer, E. A. Lacey, S. Pace, J. K. Parrish, H. M. Swain, and S. C. Trombulak. 2017. Teaching biology in the field: importance, challenges, and solutions. BioScience **67**:558-567.

Fuller, I., S. Edmondson, D. France, D. Higgitt, and I. Ratinen. 2006. International perspectives on the effectiveness of geography fieldwork for learning. Journal of Geography in Higher Education **30**:89-101.

Giles, S., C. Jackson, and N. Stephen. 2020. Barriers to fieldwork in undergraduate geoscience degrees. Nature Reviews Earth & Environment **1**:77-78.

Gilley, B., C. Atchison, A. Feig, and A. Stokes. 2015. Impact of inclusive field trips. Nature Geoscience **8**:579-580.

Ginwright, S. A., and J. Cammarota. 2015. Teaching social justice research to undergraduate students in Puerto Rico: Using personal experiences to inform research. Equity & Excellence in Education **48**:162-177.

Gold, J. R., A. Jenkins, R. Lee, J. Monk, J. Riley, I. Shepherd, D. Unwin, and D. Hiebert. 1994. Teaching Geography in Higher Education: A Manual of Good Practice. Canadian Geographer **38**:369.

Grimberg, S. J., T. A. Langen, L. D. Compeau, and S. E. Powers. 2008. A theme-based seminar on environmental sustainability improves participant satisfaction in an undergraduate summer research program. Journal of Engineering Education **97**:95-103.

Hodder, J. 2009. What are undergraduates doing at biological field stations and marine laboratories? BioScience **59**:666-672.

Hughes, A. 2016. Exploring normative whiteness: ensuring inclusive pedagogic practice in undergraduate fieldwork teaching and learning. Journal of Geography in Higher Education **40**:460-477.

Ireland, D. T., K. E. Freeman, C. E. Winston-Proctor, K. D. DeLaine, S. McDonald Lowe, and K. M. Woodson. 2018. (Un) hidden figures: A synthesis of research examining the intersectional experiences of Black women and girls in STEM education. Review of Research in Education **42**:226-254.

Jelks, S. M. R., and A. M. Crain. 2020. Sticking with STEM: Understanding STEM Career Persistence among STEM Bachelor's Degree Holders. The Journal of Higher Education **91**:805-831

John, C. M., and S. B. Khan. 2018. Mental health in the field. Nature Geoscience **11**:618-620.

Jolley, A., B. M. Kennedy, E. Brogt, S. J. Hampton, and L. Fraser. 2018a. Are we there yet? Sense of place and the student experience on roadside and situated geology field trips. Geosphere **14**:651-667.

Jolley, A., E. Brogt, B. M. Kennedy, S. J. Hampton, and L. Fraser. 2018b. Motivation and Connection to Earth on Geology Field Trips in New Zealand: Comparing American Study Abroad Students with Local Undergraduates. Frontiers: The Interdisciplinary Journal of Study Abroad.

Jolley, A., S. J. Hampton, E. Brogt, B. M. Kennedy, L. Fraser, and A. Knox. 2019. Student field experiences: designing for different instructors and variable weather. Journal of Geography in Higher Education **43**:71-95.

Jolley, A., B. M. Kennedy, E. Brogt, S. J. Hampton, and L. Fraser. 2018b. Are we there yet? Sense of place and the student experience on roadside and situated geology field trips. Geosphere **14**:651-667.

Judson, E. 2012. Learning about bones at a science museum: examining the alternate hypotheses of ceiling effect and prior knowledge. Instructional Science **40**:957-973.

Kingsbury, C. G., E. C. Sibert, Z. Killingback, and C. L. Atchison. 2020. "Nothing about us without us:" The perspectives of autistic geoscientists on inclusive instructional practices in geoscience education. Journal of Geoscience Education:1-9.

Klemow, K., A. Berkowitz, C. Cid, and G. Middendorf. 2019. Improving ecological education through a four-dimensional framework. Frontiers in Ecology and the Environment **17**:71-71.

Kortz, K. M., D. Cardace, and B. Savage. 2020. Affective factors during field research that influence intention to persist in the geosciences. Journal of Geoscience Education **68**:133-151.

Kvale, S., and S. Brinkman. 2009. Interview quality. Interviews: Learning the craft of qualitative research interviewing:161-175.

LaDue, N. D., and H. A. Pacheco. 2013. Critical experiences for field geologists: Emergent themes in interest development. Journal of Geoscience Education **61**:428-436.

Lonergan, N., and L. W. Andresen. 1988. Field-based education: Some theoretical considerations. Higher education research and development **7**:63-77.

Malm, R. H., L. M. Madsen, and A. M. Lundmark. 2020. Students' negotiations of belonging in geoscience: experiences of faculty–student interactions when entering university. Journal of Geography in Higher Education **44**:532-549.

Marín-Spiotta, E., R. T. Barnes, A. A. Berhe, M. G. Hastings, A. Mattheis, B. Schneider, and B. M. Williams. 2020. Hostile climates are barriers to diversifying the geosciences. Advances in Geosciences **53**:117-127.

Maskall, J., and A. Stokes. 2008. Designing effective fieldwork for the environmental and natural sciences. Higher Education Academy Subject Centre for Geography, Earth and ….

Milfont, T. L., and J. Duckitt. 2010. The environmental attitudes inventory: A valid and reliable measure to assess the structure of environmental attitudes. Journal of environmental psychology **30**:80-94.

Mogk, D. W., and C. Goodwin. 2012. Learning in the field: Synthesis of research on thinking and learning in the geosciences. Geological Society of America Special Papers **486**:131-163.

Morales, N., K. Bisbee O'Connell, S. McNulty, A. Berkowitz, G. Bowser, M. Giamellaro, and M. N. Miriti. 2020. Promoting inclusion in ecological field experiences: Examining and overcoming barriers to a professional rite of passage. The Bulletin of the Ecological Society of America:e01742.

Nairn, K. 1996. Parties on geography fieldtrips: embodied fieldwork? Women's Studies Journal **12**:86.

Nairn, K. 1999. Embodied fieldwork. Journal of geography **98**:272-282.

Nairn, K. 2003. What Has the Geography of Sleeping Arrangements Got to Do with the Geography of Our Teaching Spaces? Gender, Place & Culture **10**:67-81.

National Academies of Sciences, Engineering, and Medicine (NASEM). Undergraduate research experiences for STEM students: Successes, challenges, and opportunities. National Academies Press, 2017.

National Research Council (NRC). 2014. Enhancing the value and sustainability of field stations and marine laboratories in the 21st century. National Academies Press.

National Research Council (NRC). BIO2010: Transforming undergraduate education for future research biologists. National Academies Press, 2003.Cervato, C., and R. Frodeman. 2012. The significance of geologic time: Cultural, educational, and economic

O'Brien, L. T., H. L. Bart, and D. M. Garcia. 2020. Why are there so few ethnic minorities in ecology and evolutionary biology? Challenges to inclusion and the role of sense of belonging. Social Psychology of Education **23**:449-477.

O'Connell, K., K. Hoke, and R. Nilson. 2018. Landscape Study of Undergraduate Field Experiences: Report from the Field on the Design, Outcomes, and Assessment of Undergraduate Field Experiences.

O'Connell, K., K. Hoke, A. Berkowitz, J. Branchaw, and M. Storksdieck. 2020. Undergraduate learning in the field: Designing experiences, assessing outcomes, and exploring future opportunities. Journal of Geoscience Education:1-14.

O'Connell, K., K. Hoke, M. Giamellaro, A.R. Berkowitz, J. Branchaw. 2021. In press. Designing and Studying Student-Centered Undergraduate Field Experiences: The UFERN Model. BioScience.

O'Connor, C., and H. Joffe. 2020. Intercoder reliability in qualitative research: debates and practical guidelines. International Journal of Qualitative Methods **19**:1609406919899220.

Palmer, J. A., and J. Suggate. 1996. Influences and experiences affecting the pro-environmental behaviour of educators. Environmental Education Research **2**:109-121.

Patrick, A. O. 2010. Effects of field studies on learning outcome in Biology. Journal of Human Ecology **31**:171-177.

Patton, M. Q. 2008. Utilization-focused evaluation. Sage publications.

Peacock, J., R. Mewis, and D. Rooney. 2018. The use of campus based field teaching to provide an authentic experience to all students. Journal of Geography in Higher Education **42**:531-539.

Peasland, E. L., D. C. Henri, L. J. Morrell, and G. W. Scott. 2019. The influence of fieldwork design on student perceptions of skills development during field courses. International Journal of Science Education **41**:2369-2388.

Pender, M., D. E. Marcotte, M. R. S. Domingo, and K. I. Maton. 2010. The STEM pipeline: The role of summer research experience in minority students' Ph. D. aspirations. education policy analysis archives **18**:1.

Petcovic, H. L., P. M. McNeal, S. C. Nyarko, and M. H. Doorlag. 2019. "How did you learn to map?" A model for describing influential learning experiences in geologic mapping. Journal of Geoscience Education:1-17.

Petcovic, H. L., A. Stokes, and J. L. Caulkins. 2014. Geoscientists' perceptions of the value of undergraduate field education. GSA Today **24**:4-10.

Pouselt, J.R. 2020. Equity in Science Education: Representation, Culture, and the Dynamics of Change in Graduate Education. Stanford University Press.

Riggs, E. M. 2005. Field-based education and indigenous knowledge: Essential components of geoscience education for Native American communities. Science Education **89**:296-313.

Rudwick, M. 1996. Geological Travel and Theoretical Innovation: The Role of Liminal Experience. Social studies of science **26**:143-159.

Ryan, F., M. Coughlan, and P. Cronin. 2009. Interviewing in qualitative research: The one-to-one interview. International Journal of Therapy and Rehabilitation **16**:309-314.

Saldaña, J. 2011. Fundamentals of qualitative research. OUP USA.

Saldaña, J. 2015. The coding manual for qualitative researchers. Sage.

Scott, G. W., R. Goulder, P. Wheeler, L. J. Scott, M. L. Tobin, and S. Marsham. 2012. The value of fieldwork in life and environmental sciences in the context of higher education: A case study in learning about biodiversity. Journal of Science Education and Technology **21**:11-21.

Scott, G. W., S. Humphries, and D. C. Henri. 2019. Expectation, motivation, engagement and ownership: using student reflections in the conative and affective domains to enhance residential field courses. Journal of Geography in Higher Education **43**:280-298.

Semken, S. 2005. Sense of place and place-based introductory geoscience teaching for American Indian and Alaska Native undergraduates. Journal of Geoscience Education **53**:149-157.

Semken, S., E. G. Ward, S. Moosavi, and P. W. Chinn. 2017. Place-based education in geoscience: Theory, research, practice, and assessment. Journal of Geoscience Education **65**:542-562.

Shackman, G. 2008. What is program evaluation? A Beginners Guide. The **5**.

Shortlidge, E. E., and S. E. Brownell. 2016. How to Assess Your CURE: A Practical Guide for Instructors of Course-Based Undergraduate Research Experiences†. Journal of microbiology & biology education **17**:399-408.

Simm, D., and A. Marvell. 2015. Gaining a “sense of place”: Students’ affective experiences of place leading to transformative learning on international fieldwork. Journal of Geography in Higher Education **39**:595-616.

Smith, D. 2004. Issues and trends in higher education biology fieldwork. Journal of Biological Education **39**:6-10.

Stokes, A., and A. P. Boyle. 2009. The undergraduate geoscience fieldwork experience: Influencing factors and implications for learning. Field geology education: Historical perspectives and modern approaches **461**:291.

Stokes, A., A. D. Feig, C. L. Atchison, and B. Gilley. 2019. Making geoscience fieldwork inclusive and accessible for students with disabilities. Geosphere **15**:1809-1825.

Streule, M., and L. Craig. 2016. Social learning theories—An important design consideration for geoscience fieldwork. Journal of Geoscience Education **64**:101-107.

Swing, K., E. Braker, P. Fiedler, I. Billick, C. Lorentz, and D. Wagner. 2021. Growing Threats to the Scientific and Educational Legacies of Research Stations and Field Courses. BioScience.

Taylor-Powell, E., and E. Henert. 2008. Developing a logic model: Teaching and training guide. Benefits **3**:1-118.

Thomas, R. C., and S. Roberts. 2009. Experience One: Teaching the geoscience curriculum in the field using experiential immersion learning. Field geology education: Historical perspectives and modern approaches **461**:65.

Towne, L., and R. J. Shavelson. 2002. Scientific research in education. National Academy Press Publications Sales Office.

Tripp, B., S. A. Voronoff, and E. E. Shortlidge. 2020. Crossing Boundaries: Steps Toward Measuring Undergraduates' Interdisciplinary Science Understanding. CBE—Life Sciences Education **19**:ar8.

Van Der Hoeven Kraft, K. J., L. Srogi, J. Husman, S. Semken, and M. Fuhrman. 2011. Engaging students to learn through the affective domain: A new framework for teaching in the geosciences. Journal of Geoscience Education **59**:71-84.

Ward, E. G., D. Dalbotten, N. B. Watts, and A. Berthelote. 2018. Using place-based, community-inspired research to broaden participation in the geosciences. GSA Today **28**.

Whitmeyer, S. J., D. W. Mogk, and E. J. Pyle. 2009a. Field geology education: Historical perspectives and modern approaches. Geological Society of America.

Whitmeyer, S. J., D. W. Mogk, and E. J. Pyle. 2009b. An introduction to historical perspectives on and modern approaches to field geology education. Geological Society of America Special Papers **461**:vii-ix.

Wieman, C. 2012. Applying new research to improve science education. Issues in science and technology **29**:25-32.

Wiggins, G. 1998. Educative Assessment. Designing Assessments to Inform and Improve Student Performance. ERIC.

Wiggins, G. P., J. McTighe, L. J. Kiernan, and F. Frost. 1998. Understanding by design. Association for Supervision and Curriculum Development Alexandria, VA.

Williams, D. R., and J. J. Vaske. 2003. The measurement of place attachment: Validity and generalizability of a psychometric approach. Forest science **49**:830-840.

Wilson, A. E., J. L. Pollock, I. Billick, C. Domingo, E. G. Fernandez-Figueroa, E. S. Nagy, T. D. Steury, and A. Summers. 2018. Assessing science training programs: Structured undergraduate research programs make a difference. BioScience **68**:529-534.

Wilson, E. O. 1982. The importance of biological field stations. BioScience **32**:320-320.

Zavaleta, E. S., R. S. Beltran, and A. L. Borker. 2020. How Field Courses Propel Inclusion and Collective Excellence. Trends in ecology & evolution **35**:953-956.

Table Legend

Table 1. Intended student outcomes and aligned assessment tool examples. The intended student outcomes were first identified from the UFERN landscape study (O'Connell et al. 2020) and by participants at the 2018 UFERN Network Meeting at Kellogg Biological Station, April 30 – May 2, 2018. The authors of this essay then refined the list by removing those outcomes that were either duplicated, irrelevant, not measurable, or linked to very specific contexts (not field universal). Each outcome is grouped according to a primary aim defined in the table below. The table organizes published assessment tools that fall under each primary aim category and that are applicable for use in undergraduate field education experiences. This table was designed to help practioners identify instruments that align with the intended student outcomes they have identified for their field experiences. The primary aims are categories that the authors have defined to link outcomes with assessments using language that is accessible to the practitioner. The aim categories do not necessarily represent specific constructs or scales for individual assessments. The structure of the table follows that designed by Shortlidge and Brownell (2016).

Table 1. The intended student outcomes were first identified from the UFERN landscape study (O’Connell et al. 2020) and by participants at the 2018 UFERN Network Meeting at Kellogg Biological Station, April 30 – May 2, 2018. The authors of this essay then refined the list by removing those outcomes that were either duplicated, irrelevant, not measurable, or linked to very specific contexts (not field universal). Each outcome is grouped according to a primary aim defined in the table below. The table organizes published assessment tools that fall under each primary aim category and that are applicable for use in undergraduate field education experiences. This table was designed to help practioners identify instruments that align with the intended student outcomes they have identified for their field experiences. The primary aims are categories that the authors have defined to link outcomes with assessments using language that is accessible to the practitioner. The aim categories do not necessarily represent specific constructs or scales for individual assessments. The structure of the table follows that designed by Shortlidge and Brownell (2016).

Primary Aim	**Example Student Outcomes**	**Example Assessment Tools for Measuring Aim**	**Measurement details (# of items, item type, time to administer)**	**Population(s) tested**	**Ease of Analysis**	**Original Reference**
Broader Relevance - development of awareness and connection beyond the context of the field	• Increased sense of connection to local/community problems or issues • Increased sense of connection to large-scale	**Perceived Cohesion Scale (PCS)**	6 items, Likert	Multiple ages & populations	Easy	Bollen, K. A., and R. H. Hoyle. 1990. Perceived cohesion: a conceptual and empirical examination. Soc. Forces 69(2):479–504.

experience	problems or issues - Development as informed citizens					
Connection to Place - relationships between people and the field environment	- Increased stewardship intention or behaviors - Increased respect or care for the environment - Stronger connections to place	**Environmental Orientations (ECO)**	16 items, Likert	Ages 6 – 13	Easy	Larson, L. R., Green, G. T., & Castleberry, S. B. (2011). Construction and Validation of an Instrument to Measure Environmental Orientations in a Diverse Group of Children. Environment and Behavior, 43(1), 72–89. https://doi.org/10.1177/0013916509345212
		Environmental Attitudes Inventory (EAI)	24 or 72 items, Likert	Multiple ages & populations	Easy	Milfont, T.L., and J. Duckitt. (2010). The environmental attitudes inventory: a valid and reliable measure to assess the structure of enviornmental attitudes. J. Envrion. Psychol. 30: 80-94.
		Place Attachment Inventory (PAI)	15 items, Likert	Multiple ages & populations	Easy	Williams, D.R., and Vaske, J.J., 2003, The measurement of place attachment: validity and generalizability of a psychometric approach: Forest Science, v. 49, p. 830-840.
		Place Meaning Questionnaire	30 items, Likert	Multiple ages & populations	Easy	Young, M., 1999, The social construction of tourist places: Australian Geographer, v. 30,

		(PMQ)				p. 373-389, doi:10.1080/00049189993648.
		Place Meaning Scale-Marine Environments (PMS-ME)	34 items, Likert	Tourist industry representatives; resource managers; and recreational visitors	Easy	Wynveen, C. J., & Kyle, G. T. (2015). A place meaning scale for tropical marine settings. Environmental management, 55(1), 128-142.
		New Ecological Paradigm Scale (NEP)	15 items, Likert	Multiple ages & populations	Easy	Dunlap, R., K. Liere, A. Mertig, and R.E. Jones. 2000. Measuring endorsement of the new ecological paradigm: a revised NEP scale. J. Soc. Iss. 56: 425-442.
Nature of Science - Understanding of the process of science and how scientific knowledge is generated	• Increased awareness of scientific ethics • Stronger sense of what life as a scientist is like • Increased knowledge of the nature of science • Increased proficiency in	**Colorado learning attitudes about science survey - biology (CLASS-Bio)**	31 items, Likert	Undergraduate students (University of Colorado and University of British Columbia)	Moderate	Semsar, K., Knight, J.K., Birol, G., and Smith, M.K. (2011). The Colorado Learning Attitudes about Science Survey (CLASS) for use in biology. CBE-Life Sciences Education, 10, 268-278.
		Views on the Nature of Science (VNOS-C)	Open-ended, 45-60 minutes	Multiple ages & populations	Hard (requires inter-rater review of	Lederman, N. G., F. Abd-El-Khalick, R. L. Bell, and R. S. Schwartz. 2002. Views of nature of science questionnaire: toward valid and meaningful assessment of learners’ conceptions of nature of science. J.

	general research practices				answers)	Res. Sci. Teach. 39:497–521.
		Biological Experimental Design Concept Inventory (BEDCI)	14 items, multiple choice, 18 minutes	Undergraduate students (University of British Columbia)	Easy	Deane, T., K. Nomme, E. Jeffery, C. Pollock, and G. Birol. 2014. Development of the biological experimental design concept inventory (BEDCI). CBE Life Sci. Educ. 13:540–551.
		Expanded Experimental Design Ability Test (E-EDAT)	Open-ended	Undergraduate students (University of Washington)	Moderate (Rubric)	S. E. Brownell, M.P. Wenderoth, R. Theobald, N. Okoroafor, M. Koval, S.Freeman, C. L. Walcher-Chevillet, A.J. Crowe, How Students Think about Experimental Design: Novel Conceptions Revealed by in-Class Activities, BioScience, Volume 64, Issue 2, February 2014, Pages 125–137, https://doi.org/10.1093/biosci/bit016
		Experimental Design Ability Test (EDAT)	Open-ended, 10-12 minutes	Undergraduate students, Introductory class (Bowling Green State)	Moderate (Rubric)	Sirum, K., and J. Humburg. 2011. The experimental design ability test (EDAT). Bioscene J. Coll. Biol. Teach. 37:8–16
		The Rubric for Science Writing	Open ended	Undergraduates students and Graduate	Moderate (Rubric)	Timmerman, B. E C., D. C. Strickland, R.L. Johnson, and J. R. Payne. 2011. Development of a ‘universal’ rubric for

				teaching assistants (University of Southern California)		assessing undergraduates’ scientific reasoning skills using scientific writing. Assess. Eval. Higher Educ. 36:509–547.
		Test of Scientific Literacy Skills (TOSLS)	Multiple Choice, 30 minutes	Multiple populations	Easy	Gormally, C., P. Brickman, and M. Lutz. 2012. Developing a test of scientific literacy skills (TOSLS): measuring undergraduates’ evaluation of scientific information and arguments. CBE Life Sci. Educ. 11:364–377.
		Student perceptions about earth science survey (SPESS)	29 items, Likert	Undergraduate students in earth and ocean sciences (University of British Columbia)	Moderate	Jolley, A., Lane, E., Kennedy, B., and Frappé-Sénéclauze, T. 2012. SPESS: a new instrument for measuring student perceptions in earth and ocean science. Journal of Geoscience Education, 60(1):83-91.
		Entering Research Learning Assessment (ERLA)	53 items, with 47 item optional paired assessment for	Multiple populations of undergraduate and graduate trainees	Moderate (scoring guide)	Butz, A. R., & Branchaw, J. L. (2020). Entering Research Learning Assessment (ERLA): Validity evidence for an instrument to measure undergraduate and graduate research trainee development. CBE – Life Sciences Education, 19(2) https://doi.org/10.1187/cbe.19-07-0146

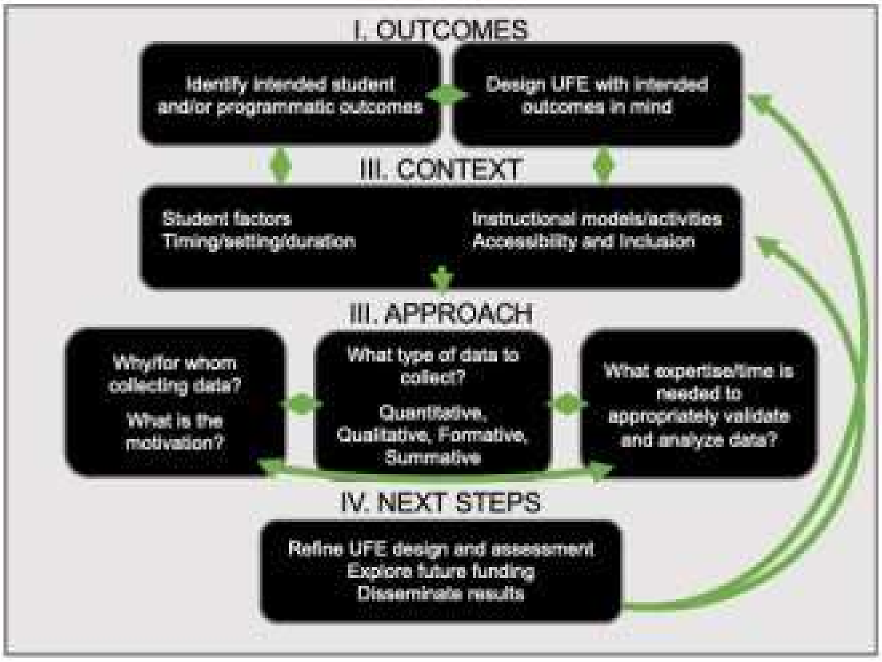
			mentors to assess trainee gains			
		Views about Science Survey (VASS)	30 items, Likert	8th-undergraduate students	Easy	Halloun, Ibrahim. (2001). Student Views about Science: A Comparative Survey. Beirut: Phoenix Series / Educational Research Center, Lebanese University.
Personal Gains - cognitive (e.g. content knowledge), behavioral (e.g. skills), and affective characteristics (e.g. comfort, confidence, self-efficacy) gained through field experience	• Ability to live and work in primitive or adverse camping conditions • Development of or increased “Grit” (perseverance through tough situation) • Increased content knowledge • Increased interest in the	**Grit Scale (GRIT)**	8 or 12 items, Likert	Multiple populations	Easy	Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. Journal of Personality and Social Psychology, 92(6), 1087-1101.
		Climate change concept inventory	21 items, Likert	Undergraduate students	Easy	Libarkin, J. C., Gold, A. U., Harris, S. E., McNeal, K. S., & Bowles, R. P. (2018). A new, valid measure of climate change understanding: associations with risk perception. Climatic Change, 150(3-4), 403-416.
		Geoscience concept inventory (GCI)	select 15 question subset from 73 total questions, Multiple choice	Undergraduate students	Easy	Libarkin, J.C., Anderson, S.W., (2006). The Geoscience Concept Inventory: Application of Rasch Analysis to Concept Inventory Development in Higher Education: in Applications of Rasch Measurement in Science Education, ed. X. Liu and W. Boone: JAM Publishers, p. 45-73

	topic of field course • More refined career goals • Improved discipline-specific skills • Development of outdoor skills • Increased confidence in physical fitness	**National Survey of Student Engagement (NSSE)***	70 items, Likert	Multiple populations	Easy	Kuh, G. D. 2009. The national survey of student engagement: conceptual and empirical foundations. New Direct. Inst. Res. 2009:5–20.
		Landscape identification and formation timescales (LIFT)	12 items, Multiple choice	Undergraduate students in earth and ocean sciences (University of British Columbia)	Easy	Jolley, A., Jones, F., and Harris, S. 2013. Measuring student knowledge of landscapes and their formation timespans. Journal of Geoscience Education, 61(2):240-251.
		Psychological Sense of School Membership (Class Belonging/School Belonging)	18 items, Likert	Middle school and undergraduate students	Easy	Goodenow, C. (1993). The psychological sense of school membership among adolescents: Scale development and educational correlates. Psychology in the Schools, 30, 79-90.
Personal Connections to Science Context - affective characteristics	• Greater sense of belonging in the scientific community • Increased value for the	**Common Instrument Suite (CIS)***	10 items, Likert	Grades 4 and above	Easy	https://www.thepearinstitute.org/common-instrument-suite
		Motivated strategies for learning	81 statements, Likert		Easy	Pintrich, R. R., & DeGroot, E. V. (1990). Motivational and self-regulated learning components of classroom academic

such as comfort, confidence, self-efficacy in science more broadly	interdisciplinary nature of science • Increased interest in a general science career • Increased interest in a field-based science career • Increased scientific self-efficacy	**questionnaire (MSLQ)**				performance, Journal of Educational Psychology, 82, 33-40.
		Science Interest Survey (SIS)	21 items, Likert	Middle and high school grade children from varying ethnic backgrounds	Easy	Lamb, R.L., Annetta, L., Meldrum, J. et al. MEASURING SCIENCE INTEREST: RASCH VALIDATION OF THE SCIENCE INTEREST SURVEY. Int J of Sci and Math Educ 10, 643–668 (2012). https://doi.org/10.1007/s10763-011-9314-z
		Career Decision Making Survey - Self Authorship (CDMS-SA)	18 items, Likert	Multiple populations	Easy	Creamer, E. G., M. B. Magolda, and J. Yue. 2010. Preliminary evidence of the reliability and validity of a quantitative measure of self-authorship. J. Coll. Student Devt. 51:550–562
		Research on the Integrated Science Curriculum (RISC)	Likert, adaptable		Easy	https://www.grinnell.edu/academics/centers-programs/ctla/assessment/risc
		Student Assessment of Learning Gains (SALG)	5 item, Likert	College students (CSU-Fullerton)	Easy	Student Perspectives on Curricular Change: Lessons from an Undergraduate Lower-Division Biology Core Merri Lynn Casem CBE—Life Sciences Education 2006 5:1, 65-75

	Science Motivation Questionnaire II (SMQII)	25 item, Likert	College students (University of Georgia)	Easy	Glynn, S. M., P. Brickman, N. Armstrong, and G. Taasoobshirazi. 2011. Science motivation questionnaire II: validation with science majors and nonscience majors. J. Res. Sci. Teach. 48:1159–1176.
	Survey of Undergraduate Research Experiences (SURE)	15 minute, Likert		Easy	Lopatto, D. 2004. Survey of undergraduate research experiences (SURE): first findings. Cell Biol. Educ. 3:270–277.
	Undergraduate Student Self-Assessment Instrument (URSSA)	Likert, adaptable	Multiple undergraduates - geared towards URE but mostly applicable	Easy	The Undergraduate Research Student Self-Assessment (URSSA): Validation for Use in Program Evaluation Timothy J. Weston and Sandra L. Laursen CBE—Life Sciences Education 2015 14:3
	STEM Self-efficacy (STEM-SE)	29 items including demographic questions, Likert	Undergraduate students but with emphasis on historically underrepresented racial/ethnic groups in science majors	Easy	Byars-Winston A, Rogers J, Branchaw J, Pribbenow, Hanke R, Pfund C. (2016). New measures assessing predictors of academic persistence for historically underrepresented racial/ethnic undergraduates in science. CBE Life Sciences Education, 3ar32.

				engaged in research experiences		
		STEM Career Interest Survey (STEM-CIS)	44 items, Likert	Middle school students (grades 6–8) who primarily were in rural, high-poverty districts in the southeastern USA	Easy	Kier M, Blanchard M, Osborne J, Albert J. (2014). The development of the STEM career interest survey (STEM-CIS). Research in Science Education 44:461-481.
Transferable Skills - skills that can be applied to contexts outside of science	• Improved communication skills • Improved collaboration skills • Improved problem-solving skills • Improved critical thinking skills	**Critical Thinking Assessment Test (CAT)***	15 items, Open-ended	Multiple populations	Moderate (scoring guide)	Stein, B., A. Haynes, M. Redding, T. Ennis, and M.Cecil. (2007). Assessing critical thinking in STEM and beyond, p 79–82. In: Innovations in e-learning, instruction technology, assessment, and engineering education. Springer, Netherlands
		California Critical Thinking Skills Test (CCTST)*	45 minutes, Multiple choice	Undergraduate students (CSU Fullerton)	Easy	Facione, P. A. 1991. Using the California Critical Thinking Skills Test in Research, Evaluation, and Assessment. [Online.]



ece3_8241_f1.jpg

ece3_8241_f2.eps