Routledge
Taylor & Francis Group

Environmental Education Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ceer20

# Development of the inventory of biotic climate literacy (IBCL)

**Emily A. Holt, Jessica Duke, Ryan Dunk & Krystal Hinerman**

**To cite this article:** Emily A. Holt, Jessica Duke, Ryan Dunk & Krystal Hinerman (09 Feb 2024): Development of the inventory of biotic climate literacy (IBCL), Environmental Education Research, DOI: 10.1080/13504622.2024.2314047

**To link to this article:** https://doi.org/10.1080/13504622.2024.2314047



Published online: 09 Feb 2024.

ENVIRONMENTAL EDUCATION RESEARCH
https://doi.org/10.1080/13504622.2024.2314047

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

Check for updates

# Development of the inventory of biotic climate literacy (IBCL)

Emily A. Holt[a], Jessica Duke[a], Ryan Dunk[a,b] and Krystal Hinerman[c]

[a]Department of Biological Sciences, University of Northern CO, Greeley, CO, USA; [b]Department of Biological Sciences, Auburn University, Auburn, AL, USA; [c]Department of Educational Leadership, Lamar University, Beaumont, TX, USA

**ABSTRACT**

Student understanding of climate change is an active and growing area of research, but little research has documented undergraduate students' knowledge about the biotic impacts of climate change. Here, we address this literature gap by presenting the Inventory of Biotic Climate Literacy (IBCL), a concept inventory developed to assess undergraduate biology student knowledge of how climate change impacts living things. We developed the IBCL through literature review, student and expert interviews, student field tests, and expert review. We implemented two large nationwide field tests and conducted multiple psychometric analyses on these datasets. These analyses resulted in a final tool of 30 items measuring 16 constructs related to the biotic impacts of climate change. We discovered that the final IBCL does not represent a single, simple construct but rather the complicated and interactive concepts that comprise this topic. We suggest that sum scores are still a valuable measure, as certain groups (upperclassmen and politically liberal individuals) scored significantly higher. We also found value in analyzing individual student performance on the IBCL by developing student profiles. The IBCL represents an important tool in assessing student understanding of the complex and growing problem of climate change and its impact on the living world.

**ARTICLE HISTORY**
Received 21 September 2023
Accepted 28 January 2024

**KEYWORDS**
Concept inventory; climate change; biology; ecology; undergraduate; SDG: 4 quality education

## Introduction

Global climate change is often referred to as a 'wicked problem' because it is ill-defined, understanding and resolving it are co-dependent, the hunt for its solutions are endless, and those solutions may cause unforeseen repercussions (Rittel and Webber 1974). More recently, climate change was elevated to a 'super-wicked problem' due to its time sensitivity and lacking central authority in managing the issue as a global problem (Cross and Congreve 2021). The gravity and complexity of climate change centers it as a topic for societal concern. However, its biological implications (Grimm et al. 2013; Walther et al. 2002) and human impacts (McMichael 2013; Patz et al. 2005) are some of its most worrying outcomes.

Climate change affects every continent on earth and every type of ecosystem, aquatic and terrestrial (IPCC, 2022). Climate effects on species are numerous and wide-ranging, including

**CONTACT** Emily A Holt emily.holt@unco.edu Department of Biological Sciences, University of Northern CO, Ross Hall 2480, Campus Box 92, Greeley, CO, USA, 80369.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/13504622.2024.2314047.

reductions in growth (Huang et al. 2021), consequences for reproduction (Parratt et al. 2021), and reduced survivability, especially when in combination with other effects (Srinivasan and Wilcove 2021). More dangerously, species-level effects compound and interact to result in a multitude of community and ecosystem-level responses. For example, climate change opens channels for concerns that were only background occurrences in the past (e.g. viral transmission and spillover, Carlson et al. 2022) and can alter ecosystem function (e.g. wetlands from sinks to sources, Salimi, Almuktar, and Scholz 2021). Due to the ubiquity and damage from climate effects on the biota, educators have been and continue to be encouraged to include climate change in science curricula (e.g. Mahaffy et al. 2017, Cervato et al. 2018, Dunk et al. 2022).

College faculty across disciplines report a responsibility to teach undergraduates about climate change because of its socio-scientific importance and its contribution to building scientific literacy (Beck, Sinatra, and Lombardi 2013), with a majority of faculty indicating that climate change should be taught in life sciences (Wise 2010). Within the field of biology, *Vision and Change* (AAAS, 2010) challenged biology undergraduate students and their educators to better understand and address global climate change. Furthermore, the Four-Dimensional Ecology Education (4DEE) Framework was established to guide educators in ecology instruction (Klemow et al. 2019), and notably, climate change is embedded as both a Core Ecological Concept and Human-Environment Interaction. Dunk et al. (2022) expanded the 4DEE to develop the Biotic Impacts of Climate Change Core Concepts ($BIC^4$) which clarifies how climate change includes multiple critical concepts that span the ecology curriculum.

These multiple educational frameworks founded in biology signal that inclusion of climate change education in our undergraduate biology curriculum is important. However, inclusion alone may be ineffective without means to assess what students are learning and understanding about climate change. To date, only three published concept inventories related to climate change exist: the Climate Change Concept Inventory (Jarrett, Ferry, and Takacs 2012), the Anthropogenic Climate Change Dissenter Inventory (Bentley, Petcovic, and Cassidy 2016), and the Greenhouse Effect Concept Inventory (Keller 2006). Unfortunately, all of these inventories stem from the earth and atmospheric sciences and focus on students' knowledge of the mechanisms causing global warming and of the abiotic systems important to climate change. Although climate change has significant biotic implications (IPCC, 2022; Grimm et al. 2013), no tool exists to assess or quantify student knowledge of climate science from the *biological perspective*. While some previous research includes survey questions related to the biotic impacts of climate change (e.g. Lambert and Bleicher 2013, McNeal, Walker, and Rutherford 2014, Hermans and Korhonen 2017), none exclusively focus on the biotic impacts of climate change. Here, we describe the development of the Inventory of Biotic Climate Literacy (IBCL) concept inventory, a tool designed to examine undergraduate student understanding of the biotic impacts of climate change. We aligned the content of this instrument with the $BIC^4$ (Dunk et al. 2022) to ensure sufficient coverage across as many of the core concepts of climate change as possible. This concept inventory is aimed at providing biology educators and researchers in biology, science, environmental, and sustainability education with a tool to measure students' conceptions of climate change impacts on biotic systems.

## Methods

Our goal was to develop a novel instrument to measure undergraduate biology students' understanding of the biotic impacts of climate change. To ensure the validity and reliability of our assessment, we used an iterative process for item development aligned with recommendations by Libarkin (2008), consisting of four main stages (Figure 1). In Stage 1, we identified topics and conceptions of the biotic impacts of climate change to determine the topical basis for the IBCL. In Stage 2, we developed items and responses for the IBCL and revised them using

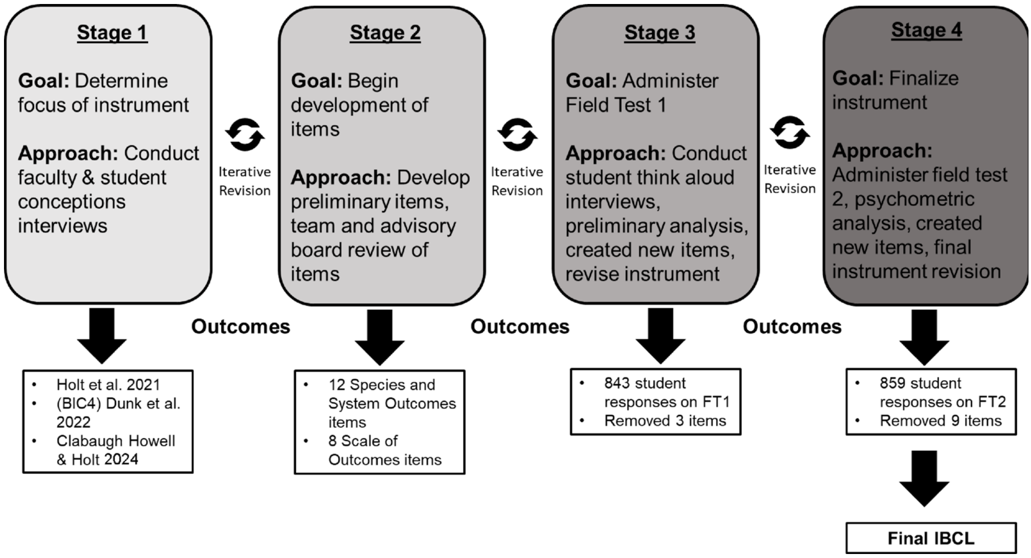



**Figure 1.** Schematic of the development process of the Inventory of biotic climate Literacy (IBCL), with four distinct stages and resulting outcomes from each stage.

expert review. In Stage 3, we revised the instrument using think-aloud student interviews, and conducted Field Test 1 for subsequent modification of the instrument. During Stage 4, we conducted a larger, more diverse Field Test 2 to broadly sample institutions across the United States and finalize the IBCL. All human participant data collection was approved by the IRB of the University of Northern Colorado (IRB #1288162), and all participants granted consent prior to completion of all surveys or interviews.

### *Stage 1 -- topic development and conceptions interviews*

During the first stage of development, our initial goal was to identify a list of topics related to the domain of the biotic impacts of climate change. The goal of this process was to collect preliminary validity evidence for our tool based on test content (*sensu* AERA, APA, & NCME et al. 2014). Through an iterative process of comparison of source data, including interviews with university faculty and college biology students and review of biology and ecology textbooks and the primary literature, we narrowed these topics into an educational framework, or the Biotic Impacts of Climate Change Core Concepts (BIC$^4$; Dunk et al. 2022). This framework centers on the consequences of climate change on the biota and its systems, irrespective of the mechanism (e.g. change in temperature, change in moisture, change in chemistry) as we understand mechanisms are interactive yet yield discrete sets of outcomes. The BIC$^4$ framework, with its three themes and seven core concepts, guided our development of conceptions interview protocols.

The second goal of Stage 1 was to identify student conceptions about the biotic impacts of climate change as a direct data source for item construction. To ensure the IBCL distractor response options and terminology aligned with our target group, we conducted a series of concept interviews with introductory and upper-division college biology students. Student interviews followed an iterative process with three rounds, with each round narrowing our focus into ideas that were directly translated into items. The collection and analysis of the first round of student interviews are described in Holt et al. (2021). Using round one interviews, we revised our interview protocol to focus only on temperature as the mechanism of climate change (i.e.

for simplicity), while also adding questions to probe student understanding of recovery patterns from climate change compared to other natural disturbances (e.g. fire). Round two student interviews were conducted with six students from two institutions, which further refined our interview protocol. The third round of conception interviews included 26 students from 10 institutions. Additionally, we used the same final protocol to interview seven biology experts (i.e. biology faculty at seven different institutions) to develop a basis for correct response options.

As part of the interview protocol, our team developed a series of illustration prompts (see Holt et al. 2021). We opted for illustration prompts for this purpose to center student ideas on the same scenario yet avoid overly taxing and complicated written descriptions. During round one and two interviews, we used illustrations of three habitats (i.e. Arctic tundra, tropical reef, and temperate forest) to prompt student ideas about how various environments would be affected by climate change. Because the greatest number of alternative conceptions arose associated with the forest scene and we felt this scene may be most familiar to undergraduates from the continental US, our team developed three additional illustrations based on this scene. These latter illustrations represented possible outcomes of the forest scene following climate change, reflecting student ideas expressed in the first sets of interviews, including minor shifts in species composition, severe fire impacts, and a catastrophic dystopia (Clabaugh Howell and Holt 2024). The three forest-based scene illustrations were the basis for the third round of conception interviews, which asked students to explain which scene represents the most probable outcome after a one degree Fahrenheit increase after 25 years. The responses from these items were primarily intended to provide a basis for items for this instrument.

### *Stage 2 – item development and revision*

Our goal for Stage 2 was to develop items using student and expert conceptions captured in our interviews from the previous stage. We adapted interview questions from our protocol to develop item stems. From our round three interview transcripts, we organized and grouped student responses thematically. The most succinct and clear of the range of responses were selected and were lightly edited for flow and clarity to represent item responses. Twelve items were developed using this process directly from the conceptions interviews, which reflected the Species and System Outcomes of the BIC$^4$ (Dunk et al. 2022). An additional eight items were drafted based on themes discussed in the interviews but did not reflect interview questions or direct participant responses. These latter eight items were created to cover the Scale Outcomes of the BIC$^4$ more fully (Dunk et al. 2022).

It is worth noting that construction of items to align with singular elements of the specific content framework (i.e. BIC$^4$; Dunk et al. 2022) proved challenging. Specifically, it was difficult to write questions about living organisms and their response to climate change without any reflection about time (the Scale Outcomes). Complete segregation of items was not possible because items about Species and Systems outcomes may have equally reflected biotic responses to any disturbance, including many that were not climate change. Likewise, items about Scale Outcomes required additional context provided by details of Species and Systems Outcomes. Although we acknowledge that cross-conceptual items are not frequently featured in conceptual assessments of student understanding, we feel that they are necessary here given the complex multidimensionality of climate change's impact on living things.

Aligned with the format of most other science concept inventories, we used a multiple-choice response format in which each item contains one correct response option and several incorrect distractor options (Dunk et al. *unpublished*; Lindell, Peak, and Foster 2007). All items included 3 to 5 response options, avoided absolutes (e.g. 'All of the above') and negatively worded statements, and ensured that all correct and distractor response options were similar in length and structure (Libarkin 2008).

We conducted an initial round ($n = 10$) of think-aloud student interviews to explore student conceptions about our items and confirm that student reasoning on items aligned with our intended reasoning (Adams and Wieman 2010). Our think-aloud interviews throughout our development process aimed to gather validity evidence based on student response processes (*sensu* AERA et al. 2014). After revising items for student readability and conceptual understandings, we then solicited feedback from our advisory board and 16 climate experts (i.e. biology researchers active in climate research) on the accuracy, readability, and BIC[4] alignment of our twenty items. Subsequently, we conducted an additional ten think-aloud interviews with biology undergraduate students to clarify that our modifications led to expected student responses. Throughout the interview and feedback process, we made iterative revisions to bring our questions in line with student thinking. Revisions in response to feedback included minor changes to wording of items as well as removal of three items and a few response options. Our instrument administered as Field Test 1 included 17 items and numerous demographic questions.

### *Stage 3 – Field Test 1*

During Stage 3, we coordinated with 21 faculty instructors from 14 different institutions nationwide for recruitment of the first field test of our 17 items. Faculty were identified through personal contacts and those who had participated in earlier stages of this project. Data were collected online *via* Qualtrics between Nov and Dec 2021 and represent post-responses as they occurred at the end of biology instruction, either introductory or upper division ecology. A full list of

**Table 1.** Institution and course demographics. Rough equivalence of samples used in the Field Test 1 and Field Test 2 sampling. Field Test 2 dataset was pruned to include only complete surveys and those in which students took 10-90 min to complete (i.e. Final dataset). Pairwise comparisons of subgroups were conducted on all groups with at least 5% of the final dataset.

| Institution Characteristic | Field Test 1 | Field Test 2 | | |
|---|---|---|---|---|
| | All Participants ($n = 1035$) | All Participants ($n = 1561$) | Final Dataset ($n = 859$) | Pairwise comparisons |
| Region | | | | |
| Midwest | 107 (10.3%) | 119 (7.6%) | 89 (10.4%) | a |
| Northeast | 50 (4.8%) | 15 (1%) | 8 (0.9%) | a |
| South | 178 (17.2%) | 426 (27.3%) | 285 (33.2%) | a |
| West | 484 (46.6%) | 677 (43.4%) | 460 (53.5%) | a |
| Missing Data | 218 (21.1%) | 324 (20.8%) | 17 (2.0%) | N/A |
| Carnegie research tier | | | | |
| R1 - Doctoral Universities, Very high research activity | 141 (13.6%) | 579 (45.6%) | 411 (47.8%) | a |
| R2 - Doctoral Universities, High research activity | 439 (42.4%) | 348 (27.4%) | 230 (26.8%) | b |
| M1 - Master's Colleges and Universities, Larger programs | 195 (18.8%) | 274 (21.6%) | 178 (20.7%) | a,b |
| M2 - Master's Colleges and Universities, Medium programs | 0 (0.0%) | 8 (0.63%) | 3 (0.3%) | N/A |
| M3 - Master's Colleges and Universities, Smaller programs | 12 (1.2%) | 7 (0.6%) | 6 (0.7%) | N/A |
| BAS - Baccalaureate Colleges, Arts & Science Focused | 30 (2.9%) | 7 (0.6%) | 5 (0.6%) | N/A |
| BDF - Baccalaureate Colleges, Diverse Fields | 0 (0.0%) | 14 (0.9%) | 9 (1.1%) | N/A |
| Other | 218 (21.1%) | 33 (2.1%) | 17 (2.0%) | N/A |
| Course Level | | | | |
| Introductory Biology Course | 559 (54.0%) | 897 (57.5%) | 596 (69.4%) | a |
| Advanced Biology Course | 279 (27.0%) | 364 (23.3%) | 260 (30.3%) | b |
| Missing Data | 197 (19.0%) | 300 (19.2%) | 3 (0.3%) | N/A |

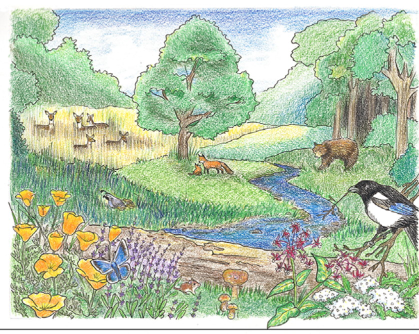
N/A: data was missing or the sample was too small (group represented <5% of the whole sample) to perform these analyses.

institutional demographics can be found in Table 1. Students were offered credit, extra credit, or no compensation, depending on instructor preference. Of the 1035 participants who took the Field Test 1 survey, 843 were complete student responses (i.e. all items were answered and they received a final score ≥ 1) and the remainder were not analyzed further. We calculated difficulty, point-biserial (i.e. item-total correlations), and discrimination on our 17 items and identified three items of concern, while the remaining 14 items demonstrated acceptable performance. We also conducted a principal components analysis (PCA) using SPSS in an attempt to identify the ways in which individual items grouped, but found no clear pattern. Rather, each item appeared to measure a nearly independent concept. This finding was a central focus of our approach to item modification.

During our item revision following Field Test 1, we created four new items aimed at increasing items that measure the unique but underrepresented BIC[4] core concepts of the spatial and temporal aspects of climate change. In addition, because the structural analysis of these field test data suggested complex factor structure, we aimed to develop additional items that closely mirrored existing questions to expand the independent constructs we detected from Field Test 1. While new items relating to core concepts already represented in the Field Test 1 version, but with new scenarios, systems, species, and scales, could further describe the biotic impacts of climate change, we felt creating mirrors would better focus on the complex, interactive concepts portrayed in the existing IBCL items. For ten of our original items and two of our new items, we created mirror items (Figure 2). These mirrors were new items that used nearly identical wording, key concepts, and distractors as existing items but changed the setting, focal species, or time period. Four of our original items, which depended on a visual scene illustration provided as part of the prompt, could

**SETUP**

**Q6.** For the following question, please use the image to the right, which shows this scene as it is currently. Imagine this image was drawn as it looked yesterday.



**ORIGINAL ITEM STEM**

**Q6a.** Over the next 25 years, the foxes in this scene will leave this area until they are no longer present in this scene. Given what you know about climate change, which of these possibilities is the most likely explanation for the disappearance of the foxes? (Form 1, 2)

**MIRRORED STEMS**

**Q6b.** Over the next 25 years, the butterflies in this scene will leave this area until they are no longer present in this scene. Given what you know about climate change, which of these possibilities is the most likely explanation for the disappearance of the butterflies? (Form 2)

**Q6c.** Over the next 25 years, the white flowers in this scene will leave this area until they are no longer present in this scene. Given what you know about climate change, which of these possibilities is the most likely explanation for the disappearance of the white flowers? (Form 1)

**RESPONSE OPTIONS**

A. Climate change will impact the area pictured in the scene, so the [foxes/butterflies/flowers] will move to an area nearby where climate change is not having any effect and survive there.
B. Climate change will impact everywhere the [foxes/butterflies/flowers] live so strongly that the [foxes/butterflies/flowers] will go extinct.
C. Climate change will impact the entire area including the scene, so the [foxes/butterflies/flowers] will move to an area nearby that now resembles their past habitat and survive there.
D. Climate change will impact the entire area including the scene, so the local [foxes/butterflies/flowers] will die out and only [foxes/butterflies/flowers] in areas far from this scene will survive.

**Figure 2.** Example item from the IBCL showing the similarity between an original item and its mirrors. The item setup (including the hand-drawn illustration by our colleague, summers scholl) was identical for all versions of this item. The top wording, in brick-color font, highlights the original stem wording that was used in both Form 1 and 2. The side-by-side wording, in mustard and blue font, represent the mirrored stems used in each form. The response choices matched each version with the appropriate wording (i.e. in this case, the varying organisms used for transfer across concepts).

not be altered as a mirror. Likewise, two of the new items discussed general patterns of the scale of climate change, for which mirrors could not be created. Our final instrument contained 41 total items, 6 were unique and unmirrored, and 12 had one or two mirrors.

### *Stage 4 – Field Test 2*

The goal of Stage 4 was to evaluate the efficacy of the IBCL with a broader audience and test new items from Stage 3. To alleviate over-burdening our participants with all 41 items, some of which were nearly identical mirrors, we prepared two test forms. Each form contained a subset of mirrored items, as well as the original item that linked response patterns across forms and all unmirrored items. This led us to 30 items on each form (Table 2), followed by demographic questions. We included randomization within our administration protocol, such that all participants were randomly assigned Form 1 or Form 2 of our instrument.

We identified faculty recruiters for the final field test through instructors who had previously participated in Field Test 1, our personal contacts, professional society listservs, and an informally constructed list of ecology instructors from 18 Historically Black Colleges and Universities (to potentially increase representation from populations largely absent from our Field Test 1 dataset). We successfully recruited 27 instructors from 20 different institutions across the United States, including 15 states. Field Test 2 was administered to biology students matriculating in 14 introductory biology courses and 17 upper division biology courses between February and May of 2022. We intentionally recruited a diversity of institution types with demographically diverse student representation to have a broader sample from which to infer. Students were offered course credit, extra credit, or no compensation, depending on instructor preference. Field Test 2 data were collected using online Qualtrics surveys with no response requirements (i.e. questions could be skipped according to our ethics protocol).

### *Statistical analysis*

All analyses were conducted using the combined sample of all courses together, and individual students were coded by course and demographic factors. The final analysis of Field Test 2 data included descriptive statistics, correlations, Classical Test Theory (CTT) and dimensional analyses (the latter aimed to collect validity evidence for internal structure, *sensu* AERA et al. 2014). First, we conducted a variety of descriptive statistics by item groups and individual items to characterize student performance across the entire assessment. Each item was scored as 1 for correct responses and 0 for incorrect responses. Overall or Total scores on the assessment were calculated by summing correct answers for each student (maximum possible score = 30). In addition to the Total score, we created three subscale scores: Climate Change; Ecology; and Blended. The Climate Change score, including two items, reflects student knowledge aligned with BIC$^4$ Scale of Outcomes theme (Dunk et al. 2022) discussing the spatial and temporal effects of climate change (Core Concepts 6, 7; Table 2). The Ecology score includes five item sets (one is mirrored and the other four are not) and strictly aligns with the BIC$^4$ Species and Systems Outcomes (Core Concepts 1–5; Table 2). The remaining nine item sets (8 mirrored and one unmirrored) sum to our Blended score, because these items aligned with, and intentionally blend, at least one Scale of Outcomes Core Concept and one Species and Systems Outcomes Core Concept (Table 2). We explored Pearson correlations among all three subscale scores, and groupings of Total scores by demographics using ANOVAs and pairwise comparisons using Tukey HSD and Bonferroni correction for multiple comparisons (alpha = 0.00625).

Second, CTT item analyses were conducted to explore item functioning (Crocker and Algina 2008). The average score of each item was used to determine the item difficulty ($P_Q$), with

**Table 2.** Content map of final 30 IBCL items (16 item sets or factors). Distribution of IBCL items across the BIC4 core concepts[1] (Dunk et al. 2022). Items aligned with CC1-5 were ecology-focused in content (Eco). Items aligned with CC6-7 were climate change-focused (clim). Items aligned with both CC1-5 (ecology-focused concepts) and CC6-7 (climate change-focused concepts) are noted as 'Blended' and span across both elements of biotic climate literacy.

| Content | Factor | Item | Form | CC1 | CC2 | CC3 | CC4a | CC4b | CC5 | CC6a | CC6b | CC7a | CC7b |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Eco | 1 | 1 | 1, 2 | X | | X | | | | | | | |
| Eco | 2 | 2a | 2 | | X | X | | | | | | | |
| Eco | | 2b | 1, 2 | | X | X | | | | | | | |
| Eco | 3 | 3 | 1, 2 | | X | | X | X | | | | | |
| Eco | 4 | 4 | 1, 2 | | | X | X | X | | | | | |
| Eco | 5 | 5 | 1, 2 | | | | X | X | | | | | |
| Blended | 6 | 6a | 1, 2 | X | | | | | | X | | | |
| Blended | | 6b | 2 | X | | | | | | X | | | |
| Blended | | 6c | 1 | X | | | | | | X | | | |
| Blended | 7 | 7a | 1, 2 | X | | | | | | | | X | X |
| Blended | | 7b | 1 | X | | | | | | | | X | X |
| Blended | | 7c | 2 | X | | | | | | | | X | X |
| Blended | 8 | 8a | 1, 2 | | X | | | | | | | X | |
| Blended | | 8b | 1 | | X | | | | | | | X | |
| Blended | | 8c | 2 | | X | | | | | | | X | |
| Blended | 9 | 9a | 1, 2 | | X | | | | | | | X | X |
| Blended | | 9b | 1 | | X | | | | | | | X | X |
| Blended | 10 | 10a | 1, 2 | | | X | | | | | | X | X |
| Blended | | 10b | 1, 2 | | | X | | | | | | X | X |
| Blended | 11 | 11a | 1, 2 | | | X | | | | | | X | X |
| Blended | | 11b | 2 | | | X | | | | | | X | X |
| Blended | | 11c | 1 | | | X | | | | | | X | X |
| Blended | 12 | 12a | 1, 2 | | | X | | | | | | X | X |
| Blended | | 12b | 2 | | | X | | | | | | X | X |
| Blended | | 12c | 1 | | | X | | | | | | X | X |
| Blended | 13 | 13 | 2 | | | | | | X | | | X | X |
| Blended | 14 | 14a | 1, 2 | | | | | | X | | | X | X |
| Blended | | 14b | 1 | | | | | | X | | | X | X |
| Clim | 15 | 15 | 1, 2 | | | | | | | X | X | | |
| Clim | 16 | 16 | 1, 2 | | | | | | | | | X | X |

[1]CC1: Species' distributions may be affected in time or space; CC2: Species' reproduction and life history traits may be affected; CC3: Species' growth and survival may be affected; CC4: Species' relationships can be affected by climate change, (a) Many types of relationships may be affected, (b) Nature of the change affecting the relationship may be direct or indirect; CC5: Biological processes can be affected; CC6: Climate change has unique spatial affects, (a) Climate change, as a disturbance, occurs on a global scale, (b) The effects of climate change are often similar across a landscape; CC7: Climate change has unique temporal effects: (a) Climate change is a continuous disturbance, (b) The observed effects of climate change occur over a long period of time (from Dunk et al. 2022).

higher values indicating lower difficulty. We calculated the point biserial correlation coefficient (PBCC) to determine test item discrimination (ranges −1 to +1). Positive PBCC values indicate that students who scored higher on the assessment also scored higher on a specific item compared to students who scored lower on the assessment, while a negative PBCC value indicates that lower scoring students scored higher on a specific item compared to those who scored higher on the assessment. The accepted threshold for PBCC is 0.3 or higher (Doran 1980).

Third, since item responses were coded as binary data (1 = correct; 0 = incorrect), we performed a Principal Components Analysis in SPSS with Varimax rotation to maximize explained variation and explore instrument cohesion. For the purposes of interpretation, factor loadings greater than 0.3 were considered viable. Unlike all other analyses conducted on all participants combined, PCAs were conducted separately on the two sample groups (Form 1, $n = 404$; Form 2, $n = 455$). We also used non-metric multidimensional scaling (NMS) in PC-ORD (v. 7.09, McCune and Mefford 2018) to further explore dimensionality and clustering of item sets. The 'slow and thorough' autopilot mode of PC-ORD sought the best solution of Form 1 data using Euclidean distances, with a maximum of 500 iterations of 50 runs of real data and 50 randomized trials.

### *Qualitative analysis*

Since the instrument was designed to capture conceptual knowledge along the broad spectrum of the BIC[4] constructs and also understand how the instrument could be used for pedagogical and curriculum improvements, we conducted an exploratory qualitative Student Performance Profile analysis to evaluate within- and between-student performance using data visualization plots. We randomly selected a set of ten high-performing participants (Total scores: 15–20) and a set of ten low-performing participants (Total scores: 5–10). The responses of these 20 students were coded by whether they selected the correct or incorrect response for every item. For mirrored item sets, responses from each item were averaged. We visualized Student Performance Profiles as a row in a table, where each item was shaded to denote their performance: unshaded was incorrect, black-shaded was correct, gray-shaded was an item with its mirror where their responses disagreed (i.e. they scored correctly on one version and incorrectly on the other version). Items that are black-shaded across multiple students are less difficult items, while unshaded items across multiple students are more difficult. Items with gray-shading suggest that they assess concepts that have low transferability across mirrors.

## Results

### *Initial analysis*

Field Test 2 resulted in 1561 participant responses. To provide a high-quality dataset that reflected our target population, we applied a participant filtering process before data analysis (Uminski and Couch 2021). We first removed all participant data in which a student did not answer all items or received a score of zero on the IBCL concept inventory ($n = 230$ and $n = 69$, respectively). We next removed participants that spent less than 10 or greater than 90 min completing the survey, in line with Couch et al. (2019), to filter respondents likely dedicating insufficient effort or overabundant time indicative of the survey being taken over more than one sitting. In total, our final data set consisted of 859 student responses split between two forms (Form 1 = 404 and Form 2 = 455). Demographic information for students in both the initial and filtered datasets, the latter of which was used for all subsequent analyses, were comparable (Table 3). Although our data cleaning approaches aligned with Enders (2022), this comparison of demographic representation suggests little impact on our sample.

### *Preliminary analysis*

The preliminary analysis of Field Test 2 data included all 41 unique items designed to capture 18 constructs. Following preliminary CTT and dimensional item analysis, we deleted items based on redundancy (one item), poor PCA loading (factor loadings <0.3; one entire item set of three items), and because the items were either too difficult and/or had poor discrimination (PBCC < 0.2; four items and one entire item set of three items). These analyses resulted in the removal of eleven poor-performing items. Removal of these items did not interfere with the integrity of our instrument because all remaining items provided adequate coverage of the BIC[4] constructs.

### *Final analysis*

The final retained dataset included 30 items and 16 constructs with 859 participants. These items and item sets maintained coverage of the BIC[4] constructs (Table 2). The mean overall score did not differ between Form 1 (F1) and Form 2 (F2) (F1:12.8 +/- 4.92 SD; F2:12.8 +/- 4.56 SD, where the possible points on both forms was 23 (Figure 3). Each form differed slightly in

**Table 3.** Student demographics. Rough equivalence of samples used in the Field Test 1 and Field Test 2 sampling. Field Test 2 dataset was pruned to include only complete surveys and those in which students took 10-90 min to complete (i.e. Final dataset). Pairwise comparisons of subgroups were conducted on all groups with at least 5% of the final dataset.

| Student Demographics | Field Test 1 | Field Test 2 | | |
|---|---|---|---|---|
| | All Participants ($n = 1035$) | All Participants ($n = 1561$) | Final Dataset ($n = 859$) | Pairwise comparisons |
| Class Standing | | | | |
| Freshman | 70 (6.8%) | 269 (17.2%) | 163 (19.0%) | a |
| Sophomore | 285 (27.5%) | 368 (23.6%) | 245 (28.5%) | a,b |
| Junior | 258 (24.9%) | 331 (21.2%) | 231 (26.9%) | a,b |
| Senior | 219 (21.2%) | 280(17.9%) | 207 (24.1%) | b |
| Post Baccalaureate | 10 (1.0%) | 15 (1.0%) | 12 (1.4%) | N/A |
| Missing Data | 193 (18.6%) | 298 (19.1%) | 1 (0.1%) | N/A |
| Gender Identity | | | | |
| Man (Cis or Trans) | 255 (24.6%) | 368 (23.6%) | 261 (30.4%) | a |
| Woman (Cis or Trans) | 562 (54.3%) | 825 (52.9%) | 548 (63.8%) | a |
| Non-binary, Genderqueer, or Gender Non-conforming | 12 (1.2%) | 32 (2.0%) | 24 (2.8%) | N/A |
| Prefer not to answer | 20 (2.1%) | 34 (2.2%) | 26 (3.0%) | N/A |
| Missing Data | 184 (17.8%) | 302 (19.3%) | 0 (0.0%) | N/A |
| Race/Ethnicity* | | | | |
| American Indian/Alaska Native | 12 (1.2%) | 27 (1.7%) | 20 (2.1%) | N/A |
| Asian | 126 (12.2%) | 158 (10.1%) | 105 (11.0%) | a |
| Black or African American | 48 (4.6%) | 80 (5.1%) | 47 (4.9%) | N/A |
| Hispanic or Latino/a/x | 334 (32.3%) | 341 (21.8%) | 234 (24.6%) | a |
| Native Hawaiian or other Pacific Islander | 3 (0.3%) | 17 (1.1%) | 8 (0.8%) | N/A |
| White | 378 (36.5%) | 721 (46.2%) | 510 (53.5%) | b |
| Prefer not to answer | 90 (8.7%) | 78 (5.0%) | 29 (3.0%) | N/A |
| Missing Data | 91 (8.8%) | 139 (8.9%) | 5 (0.5%) | N/A |
| Political Affiliation | | | | |
| Democrat | 297 (28.7%) | 413 (26.5%) | 279 (32.5%) | a |
| Green Party | 5 (0.5%) | 17 (1.1%) | 13 (1.5%) | N/A |
| Independent | 129 (12.5%) | 192 (12.3%) | 135 (15.7%) | a |
| Libertarian Party | 23 (2.2%) | 31 (2.0%) | 21 (2.4%) | N/A |
| Republican | 116 (11.2%) | 155 (9.9%) | 104 (12.1%) | a |
| Other | 32 (3.1%) | 49 (3.1%) | 35 (4.1%) | N/A |
| Unknown | 111 (10.7%) | 196 (12.6%) | 145 (16.9%) | N/A |
| I prefer not to say | 134 (12.9%) | 205 (12.1%) | 125 (14.6%) | N/A |
| Missing Data | 188 (18.2%) | 303 (19.4%) | 2 (0.2%) | N/A |
| Political Identity | | | | |
| Strongly Conservative | 22 (2.1%) | 28 (1.8%) | 19 (2.2%) | a |
| Moderately Conservative | 70 (6.8%) | 96 (6.1%) | 59 (6.9%) | |
| Slightly Conservative | 41 (4.0%) | 69 (4.4%) | 51 (5.9%) | |
| Middle of the Road | 109 (10.5%) | 168 (10.8%) | 113 (13.2%) | a |
| Slightly Progressive | 52 (5.0%) | 57 (3.7%) | 44 (5.1%) | b |
| Moderately Progressive | 147 (14.2%) | 213 (13.6%) | 162 (18.9%) | |
| Strongly Progressive | 111 (10.7%) | 210 (13.5%) | 138 (16.1%) | |
| I prefer not to say | 144 (13.9%) | 181 (11.6%) | 112 (13.0%) | N/A |
| I don't know | 152 (14.7%) | 236 (15.1%) | 159 (18.5%) | N/A |
| Missing Data | 187 (18.1%) | 303 (19.4%) | 2 (0.2%) | N/A |

*Some participants may be included in multiple race and ethnicity categories, and those who marked more than one race were not included in pairwise comparisons.

the number of items comprising subscale scores (Climate Change subscale: same 2 items on both forms; Ecology subscale: F1 = 5 items; F2 = 6 items; Blended subscale: F1 = 16 items; F2 = 15 items). Mean subscale scores for each form were Climate Change Score: F1: 1.32 +/- 0.74; F2: 1.35 +/- 0.71, Ecology Score: F1: 2.86 +/- 1.24; F2: 3.40 +/- 1.51, and Blended Scores: F1: 8.60 +/- 3.77; F2: 8.02 +/- 3.21. The subscale scores for the Blended items strongly correlated with the Climate Change items and Ecology items ($r = 0.45$ and $r = 0.47$, respectively), but these latter two subscale scores only showed moderate correlation ($r = 0.34$).

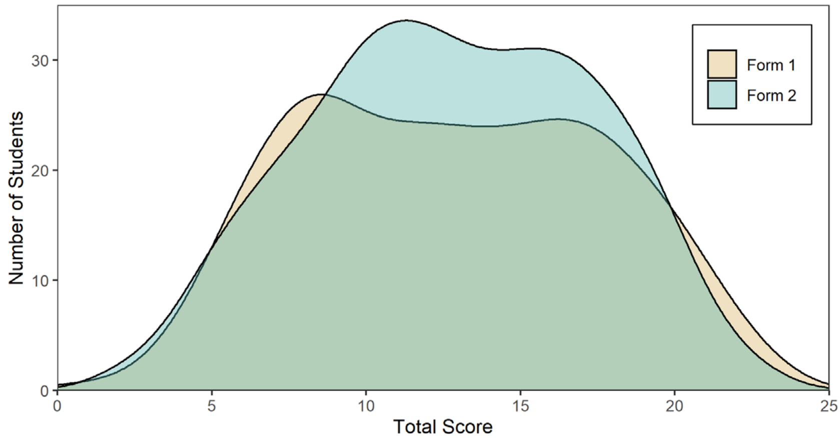



**Figure 3.** Score distribution from our Field Test 2 showing the similarity between students given Form 1 and Form 2. Graphs represent smoothed histograms of each sample population in contrasting colors.

The 30 items comprising the final version of the IBCL ranged in difficulty from 0.67 to 0.33 ($M = 0.55$), with four items being more difficult (< 0.4), ten items falling in the medium difficulty range (values between 0.40 and 0.59), and eleven items that were fairly easy (> 0.6) (Table 4). Our item discrimination values ranged between 0.25 and 0.6 ($M = 0.42$) with only three items falling below the accepted threshold of 0.3 (Table 4). PCA dimension reduction analysis indicated each item set was an independent factor and that mirrors developed to capture the same construct largely aligned with one another. In line with a multidimensional structure, instead of factors with stronger explanatory value which would represent underlying common conceptual understandings driving responses, each retained factor only explained between 3% and 5% of the total variance. As further illustration, the NMS ordination demonstrates that while mirrored items in the same set generally occupy the same item space (i.e. connected by a line, Figure 4), and items of the same subscale cluster to some degree, there is no strong separation of items along the axes shown to signal clear latent factors. Similar to analyses with Field Test 1 data, our Field Test 2 data did not exhibit unidimensionality, thus confirming the instrument is capturing 16 unique constructs. This multidimensionality precluded use of Rasch or other item response theory (IRT) methods, which are becoming increasingly popular in instrument development (Hambleton, Zenisky, and Popham 2016). Similarly, measures of internal consistency (e.g. Cronbach's alpha, KR20) are not appropriate for a multidimensional scale of this type with no grouping components (Crocker and Algina 2008).

### *Demographic comparisons*

Comparing Total IBCL scores across institutions and courses, we found no difference by region; however, students at R1 institutions and in advanced biology courses scored overall higher on the IBCL (Table 1). Exploring differences by individual demographics across our sample, we identified no significant differences by gender identity or political party affiliation. We noted that pairwise differences occurred within class standing, race/ethnicity, and political identity (Table 3). IBCL scores increased with class standing, being the highest in our sample's seniors. Students identifying with a progressive political ideology tended to have higher IBCL scores. We also found self-identified white students scored higher on the IBCL than students who self-identified as Asian or Hispanic and/or Latin(a/o/x), while we did not have enough resolution to make comparisons for any other race/ethnicity recorded.

**Table 4.** Difficulty and point Bi-serial calculations for each IBCL item from all students.

| Factor | Item | Difficulty | Point Bi-serial |
|---|---|---|---|
| 1 | 1 | 0.33 | 0.3 |
| 2 | 2a | 0.63 | 0.43 |
| | 2b | 0.62 | 0.43 |
| 3 | 3 | 0.66 | 0.31 |
| 4 | 4 | 0.58 | 0.38 |
| 5 | 5 | 0.63 | 0.41 |
| 6 | 6a | 0.58 | 0.36 |
| | 6b | 0.55 | 0.3 |
| | 6c | 0.33 | 0.25 |
| 7 | 7a | 0.58 | 0.43 |
| | 7b | 0.53 | 0.54 |
| | 7c | 0.45 | 0.25 |
| 8 | 8a | 0.63 | 0.57 |
| | 8b | 0.64 | 0.59 |
| | 8c | 0.61 | 0.6 |
| 9 | 9a | 0.55 | 0.5 |
| | 9b | 0.59 | 0.47 |
| 10 | 10a | 0.63 | 0.54 |
| | 10b | 0.61 | 0.56 |
| 11 | 11a | 0.57 | 0.38 |
| | 11b | 0.55 | 0.31 |
| | 11c | 0.6 | 0.46 |
| 12 | 12a | 0.36 | 0.34 |
| | 12b | 0.33 | 0.35 |
| | 12c | 0.42 | 0.48 |
| 13 | 13 | 0.42 | 0.28 |
| 14 | 14a | 0.53 | 0.41 |
| | 14b | 0.51 | 0.42 |
| 15 | 15 | 0.67 | 0.44 |
| 16 | 16 | 0.68 | 0.47 |

### *Student Performance Profile analysis*

To understand the complexity of the topic measured in the IBCL, manifesting as a multidimensional instrument, we qualitatively explored student performance profiles to compare performance within and between high and low performing groups of students (Figure 5). Our goal with this analysis was to identify items that helped discriminate high from low performing students, items that were generally difficult or easy, and to identify topic transferability. Mirrored items allowed us to determine the transferability of knowledge across concepts; when both the original item and its mirror were correctly answered, we assumed the concept was highly transferable (dark blue outlines, Figure 5), but when students answered one correctly and the matched item incorrectly we assumed the concept had low transferability (bright blue outlines, Figure 5). Quantifying transferability is important because 'detecting students' knowledge and misconceptions in one context will not provide evidence of competency in another context' (Nehm et al., 2012). While comparison of these mirrored items allows us to measure transfer, we are unable to understand what factors allow them to transfer knowledge or not. Item 8 (i.e. addressing CC2 - life history), where the mirrors varied only the focal organism, shows high transferability for both score groups with the majority of high-performing students answering both questions in an item set correctly, while most low-performing students answered both questions in the item set incorrectly (Figure 5). Items 6 and 7 (i.e. addressing CC1 – distribution), whose mirrors also varied the focal organism, showed low transferability across both score groups, where most students answered one of the mirrors correctly and the other incorrectly, regardless of group (gray shading, Figure 5). The transferability of items 2, 10, 11, and 12 (i.e. all addressed CC3 – growth and survival) varied by group; high-performing students tended to

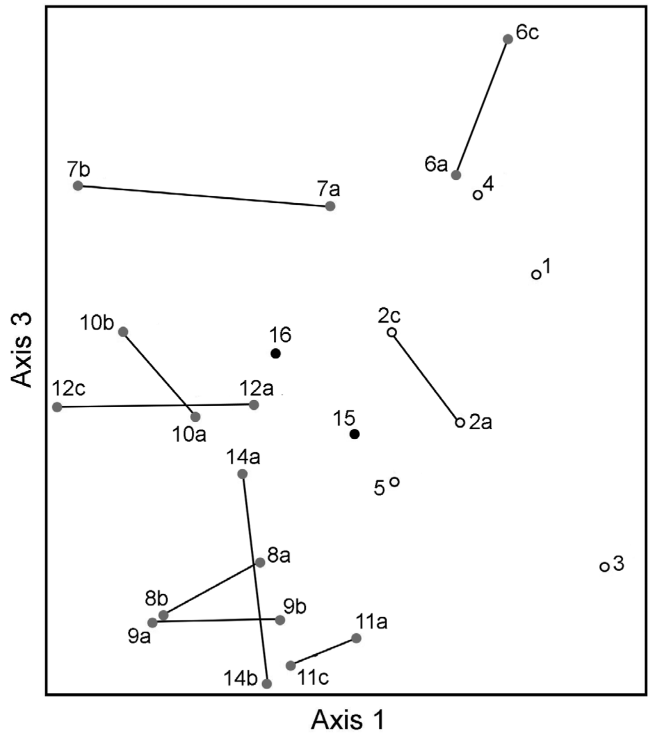



**Figure 4.** NMS ordination of Field Test 2 data. Only two of the three axes are shown to best depict the clustering of items. Circles represent each item in student space ($n = 404$ students who took Form 1). Lines connect items of the same factor (i.e. mirrors to original), and unconnected circles are items with no mirror. Open circles are 'Ecology-only' items, grey-filled circles are 'Blended' items, and black-filled circles are 'climate-only' items (see Table 2).

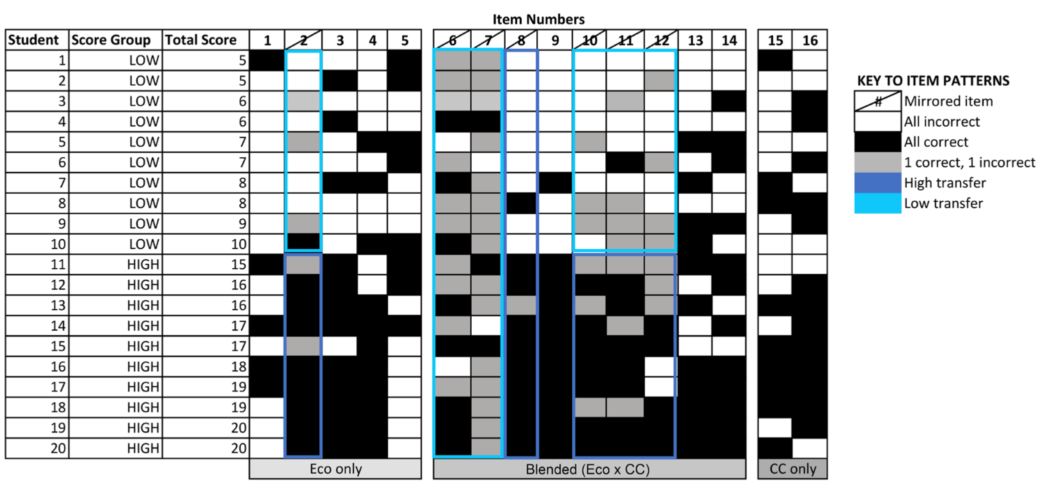



**Figure 5.** Student profile analysis including 10 low-performing students (Total score: 5-10) and 10 high performing students (Total scores: 15–20). Black-filled box indicates an item (and its mirror, if applicable) were correctly answered by that student. Unfilled boxes indicate items (and their mirrors) were incorrectly answered by that student. Gray-filled boxes indicate items that have a mirror and the student answered one correctly and the other incorrectly. Dark blue outlines denote items with high transfer and bright blue outlines indicate items with low transfer. Subscales are indicated below each item (Eco only: Ecology only items, Blended: Blended items including interacting Eco and CC topics, CC only: Climate change only items, see Table 2).

get both mirrors correct while most low-performing students answered only one of the two mirrors in an item set correctly (Figure 5).

Our student profiles also indicate that items 3 and 8–12 discriminate well within our student population (i.e. high-performing students mostly answered items correctly and low-performing students answered mostly incorrectly; Figure 5). Item 5 was the only item that did not discriminate well with our population (i.e. low-performing students answered items correctly more often than high-performing students; Figure 5). This item focused on the concept of a climatic niche to reflect findings in Holt et al. (2021) that suggested this topic could be challenging to college students in biology. Finally, our student profiles reflect trends we noted across the entire dataset, with a range of difficulty; items 1 and 5 were the highest difficulty across both groups, items 13 and 14 were moderate difficulty, and item 16 (i.e. addressing CC7 – temporal effects of climate change) was the lowest difficulty item.

The student profiles also allowed us to differentiate within-student performance across our three subscales (Climate Change, Ecology, and Blended). For example, Student 11 answered 67% of Ecology items correctly but incorrectly answered both Climate Change items, however, they answered 73% of the Blended items correctly. Alternatively, Student 8 did not answer a single Ecology item correctly but answered both Climate Change items correctly, consequently, they answered only 40% of the Blended items correctly. These patterns suggest that foundational knowledge in ecology may be critical in understanding the biotic impacts of climate change (i.e. Blended items that include scenarios that include both elements). Yet as a contrast, Student 15 only correctly identified 33% of the Ecology items but answered both Climate Change items correctly; then, they answered 87% of the Blended items correctly. These mixed results further suggest that these subscale designations may have been appropriate for development, but likely are less useful for analysis purposes and inference.

## Discussion

Our central goal was to describe the development of the Inventory of Biotic Climate Literacy (IBCL) concept inventory as a tool to examine undergraduate student understanding of the biotic impacts of climate change. Through a four-stage process including two field tests, the IBCL represents a novel instrument measuring a complex topic currently lacking detailed metrics. We assessed the validity evidence of our instrument by seeking independent feedback from experts in the field and think-aloud interviews with participants from the population of interest. We did not assess convergent validity with existing instruments; yet, propose that future work collecting IBCL data with other measures of climate change knowledge (e.g. Bentley, Petcovic, and Cassidy 2016; Jarrett, Ferry, and Takacs 2012; Keller 2006) or ecology knowledge (e.g. Hartley et al. 2011; Wilson et al. 2006) may be fruitful.

Each of our two field tests included large samples of >1000 students. The diversity in our sample suggests its generalizability across Introductory Biology and Ecology university student populations. Due to observed differences across student groups (Table 3), and in keeping with best practices for equitable instrument development, we recommend future studies continue to evaluate performance of the IBCL across varied student demographic groupings and sample diverse student populations. We expected IBCL scores to increase through a students' educational trajectory as they gain more exposure to the topic, and suggest our instrument may be useful in detecting the impact of successful interventions in future work. The mean score on Field Test 2 was 56% accuracy and only 6 individuals scored 100% accuracy. This level of overall conceptual knowledge of the biotic impacts of climate change indicates room for growth. We report only post-survey data in this study; therefore, future work integrating interventions to scaffold and build this conceptual knowledge may be able to detect changes over time and further confirm instrument soundness *via* test-rest reliability.

The 30 items in our final IBCL instrument align well with the BIC[4] (Dunk et al. 2022) and generally overlap with Ecology, Climate Change, or Blended content of both. The Blended items tended to be more difficult for our participants, likely owing to the interleaving of ecology concepts and climate literacy into a systems paradigm (Momsen et al. 2022). However, success on the Ecology only or Climate Change only items did not help predict their performance on the Blended items.

### *How to use the IBCL*

Our instrument is multidimensional and measures several overlapping constructs; therefore, users are cautioned about interpreting the Total IBCL score as a single simple construct. This total score may provide a measure of overall conceptions related to the biotic impacts of climate change that could aid group testing (e.g. within a classroom or program) to seek trends. We noted improvements in Total IBCL scores in our sample from introductory to advanced classes and by class standing (Tables 1 and 3). Therefore, we believe the IBCL will be useful to researchers and practitioners to measure overall growth in this content. Its complex instrument functioning (i.e. lacking a single salient factor) reflects the infinitely complex underlying topic; accordingly, the IBCL development and structure may represent a new type of concept inventory whose interactivity requires careful examination of student response on an item-by-item basis and less reliance on a Total score. Accordingly, we recommend that seeking patterns for smaller groups of students, similar to our Student Performance Profiles (e.g. Figure 5) may be an effective way to examine student performance and potentially to explore changes over time. These fine-scale investigations with the IBCL on an item-by-item basis may identify weak areas in student understanding of the biotic impacts of climate change and guide instructional interventions for improvement.

The IBCL will certainly be a critical tool to understand student misconceptions, which can guide curriculum development and deliberate delivery of climate content in biology contexts. Further, most of the IBCL items are scenario-based allowing for a more authentic transfer of climate change knowledge and deeper critical thinking required by students. Climate change is a complex, interdisciplinary topic that requires integration of multiple sources of information (IPCC, 2022), translation from other fields (Reis and Ballinger 2020), and evaluation of competing ideas (Heffron and Valmond 2011). Instruction and assessments that rely strictly on fact-based information do not prepare a student for the breadth of decisions or scenarios they may face impacted by climate change. We hope the IBCL can serve as an effective assessment to detect students' ability to transfer and combine information. We recommend that practitioners use all 30 items, which includes some mirrored items, when administering the IBCL. The complexity of the overall construct of the biotic impacts of climate change and the subsequent multidimensionality of the IBCL reminds practitioners that its items are measuring interactive concepts. High achievement on the IBCL signals a student who can combine concepts from several ecological areas to understand impacts from climate change that are different from other disturbances and transfer these ideas across species and scale. The full instrument is included in Supplemental/Appendix A and the answer key is available upon request.

### *Limitations*

Our study, i.e. the development of the IBCL, had limitations. During Stages 1 and 2, our list of relevant topics and domain of study were limited by those outlined in the BIC[4] (Dunk et al. 2022). We were further limited by our capacities to develop multiple unique items reflecting individual concepts from the BIC[4], both of which were mutually exclusive of other concepts yet accurately reflected different ideas still aligned with that concept. During Stages 3 and 4, our process was limited because identifying willing faculty recruiters from a diverse set of institutions representing diverse student populations was challenging. Finally, the analysis portion of our process was limited by the psychometric analyses we could use that were appropriate for the type and behavior of the data we collected. We further reiterate that our final instrument has its own limitations. Our items reflect a sampling of species, time frames, and ecosystems and their responses to climate change, yet additional scenarios are countless and the combinations therein infinite. Our focal population was American undergraduate biology students, and here the IBCL was tested on a sample of these types of participants, which limits its generalizability to other populations. Further, the final IBCL is limited in its utility due to its multidimensionality. We describe ideal ways to use this tool above, despite these challenges.

### *Next steps*

While we believe the IBCL will greatly contribute to biology education, we recognize there are many steps ahead. As mentioned above, we encourage additional testing of the IBCL, including pre-post testing and combining our instrument with other relevant, validated surveys about climate literacy. While the multidimensional nature of our instrument limits its utility to measure a single simple concept, we believe it signals that the biotic impacts of climate change are complicated and themselves multidimensional. Further, other concept inventories have been developed recently that were not unidimensional (e.g. Wasendorf et al. 2022), suggesting that multidimensionality of concept inventories may be more common than previously thought. Finally, this article introduces our new instrument and describes its development process; however, work lies ahead to more richly describe the student conceptions that underpin their responses on the IBCL.

## Acknowledgements

This work was supported by the National Science Foundation (NSF) Improving Undergraduate STEM Education (DUE-1836522). Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the NSF. We thank our many participants and experts who contributed to several stages of this project. We acknowledge the sage advice of our advisory board members (April Maskiewicz Cordero, Anna Gold, Lorna Jarrett, Sue Hyeon Paek). We are indebted to our colleague, Summers Scholl, who created the six beautifully hand-drawn illustrations used in this work to perfectly reflect our findings in the literature and vision for how to test student understanding of climate change and will live forever in the IBCL.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

Emily A. Holt http://orcid.org/0000-0002-1777-7882
Jessica Duke http://orcid.org/0000-0003-0895-0265
Krystal Hinerman http://orcid.org/0000-0003-2777-0905

## References

AAAS, American Association for the Advancement of Science. 2010. Vision and Change: A Call to Action. Washington, DC. Available at: http://visionandchange.org/files/2013/11/aaas-VISchange-web1113.pdf, accessed 27 Mar 2018.

Adams, W. K., and C. E. Wieman. 2010. "Development and Validation of Instruments to Measure Learning of Expert-like Thinking." *International Journal of Science Education* 33 (9): 1289–1312. https://doi.org/10.1080/09500693.2010.512369

AERA, APA, & NCME. 2014. *Standards for Educational and Psychological Testing: National Council on Measurement in Education*. Washington DC: American Educational Research Association.

Beck, A., G. M. Sinatra, and D. Lombardi. 2013. "Leveraging Higher-Education Instructors in the Climate Literacy Effort: Factors Related to University Faculty's Propensity to Teach Climate Change." *The International Journal of Climate Change: Impacts and Responses* 4 (4): 1–17. https://doi.org/10.18848/1835-7156/CGP/v04i04/37181

Bentley, Andrew P. K., Heather L. Petcovic, and David P. Cassidy. 2016. "Development and Validation of the Anthropogenic Climate Change Dissenter Inventory." *Environmental Education Research* 25 (6): 867–882. https://doi.org/10.1080/13504622.2016.1250150

Carlson, Colin J., Gregory F. Albery, Cory Merow, Christopher H. Trisos, Casey M. Zipfel, Evan A. Eskew, Kevin J. Olival, Noam Ross, and Shweta Bansal. 2022. "Climate Change Increases Cross-Species Viral Transmission Risk." *Nature* 607 (7919): 555–562. https://doi.org/10.1038/s41586-022-04788-w

Cervato, C., D. Charlevoix, A. Gold, and H. Kandel. 2018. "Research on Students' Conceptual Understanding of Environmental, Oceanic, Atmospheric, and Climate Science Content." In *Community Framework for Geoscience Education Research*, edited by St. John, K. National Association of Geoscience Teachers. https://doi.org/10.25885/ger_framework/3

Clabaugh Howell, K., and E. A. Holt. 2024. "Student Reasonings and Cognitive Biases in Climate Change Predictions." *Science Education* 108 (1): 86–106. https://doi.org/10.1002/sce.21826

Couch, Brian A., Christian D. Wright, Scott Freeman, Jennifer K. Knight, Katharine Semsar, Michelle K. Smith, Mindi M. Summers, Yi Zheng, Alison J. Crowe, and Sara E. Brownell. 2019. "GenBio-MAPS: A Programmatic Assessment to Measure Student Understanding of Vision and Change Core Concepts across General Biology Programs." *CBE Life Sciences Education* 18 (1): Ar1. https://doi.org/10.1187/cbe.18-07-0117

Crocker, L. M., and J. Algina. 2008. *Introduction to Classical and Modern Test Theory*. Mason, OH, US: Cengage Learning.

Cross, I. D., and A. Congreve. 2021. "Teaching (Super) Wicked Problems: Authentic Learning about Climate Change." *Journal of Geography in Higher Education* 45 (4): 491–516. https://doi.org/10.1080/03098265.2020.1849066

Doran, R. L. 1980. *Basic Measurement and Evaluation of Science Instruction*. Washington, DC: National Science Teachers Association, US.

Dunk, R., J. Sexton, K. Hinerman, and E. A. Holt. 2022. "Development of the Biotic Impacts of Climate Change Core Concepts (BIC4) Framework." *Environmental Education Research* 28 (8): 1175–1190. https://doi.org/10.1080/13504622.2022.2069682

Enders, C. K. 2022. *Applied Missing Data Analysis*. 2nd ed. New York, NY, US: The Guilford Press.

Grimm, Nancy B., F. Stuart Chapin, III, Britta Bierwagen, Patrick Gonzalez, Peter M. Groffman, Yiqi Luo, Forrest Melton, et al. 2013. "The Impacts of Climate Change on Ecosystem Structure and Function." *Frontiers in Ecology and the Environment* 11 (9): 474–482. https://doi.org/10.1890/120282

Hambleton, R. K., A. L. Zenisky, and W. J. Popham. 2016. "Criterion-Referenced Testing: Advances over 40 Years." In *Educational Measurement: From Foundations to Future*, edited by Wells, C. S. and Falkner-Bond, M, 23–37. Guilford Press, NY, US.

Hartley, L. M., B. J. Wilke, J. W. Schramm, C. D'Avanzo, and C. W. Anderson. 2011. "College Students' Understanding of the Carbon Cycle: Contrasting Principle-Based and Informal Reasoning." *BioScience* 61 (1): 65–75. https://doi.org/10.1525/bio.2011.61.1.12

Heffron, S. G., and K. Valmond. 2011. "Teaching about Global Climate Change." *The Geography Teacher* 8 (2): 91–95. https://doi.org/10.1080/19338341.2011.571154

Hermans, M., and J. Korhonen. 2017. "Ninth Graders and Climate Change: Attitudes towards Consequences, Views on Mitigation, and Predictors of Willingness to Act." *International Research in Geographical and Environmental Education* 26 (3): 223–239. https://doi.org/10.1080/10382046.2017.1330035

Holt, E. A., A. B. Heim, J. Sexton, and K. Hinerman. 2021. "Undergraduate Student Conceptions of Climate Change Impacts on Animals." *Ecosphere* 12 (8): E 03706. https://doi.org/10.1002/ecs2.3706

Huang, M., L. Ding, J. Wang, C. Ding, and J. Tao. 2021. "The Impacts of Climate Change on Fish Growth: A Summary of Conducted Studies and Current Knowledge." *Ecological Indicators* 121: 106976. https://doi.org/10.1016/j.ecolind.2020.106976

IPCC, Intergovernmental Panel on Climate Change. 2022. *Climate Change 2022: Impacts, Adaptation, and Vulnerability*. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, New York, NY, US. https://doi.org/10.1017/9781009325844

Jarrett, L., B. Ferry, and G. Takacs. 2012. "Development and Validation of a Concept Inventory for Introductory-Level Climate Change Science." *International Journal of Innovation in Science and Mathematics Education* 20 (2): 25–41.

Keller, J. M. 2006. "Development of a Concept Inventory Addressing Students' Beliefs and Reasoning Difficulties regarding the Greenhouse Effect." PhD diss., University of Arizona, Tucson, AZ, US.

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 (2): 71–71. https://doi.org/10.1002/fee.2013

Lambert, J. L., and R. E. Bleicher. 2013. "Climate Change in the Preservice Teacher's Mind." *Journal of Science Teacher Education* 24 (6): 999–1022. https://doi.org/10.1007/s10972-013-9344-1

Libarkin, J. 2008. *Concept Inventories in Higher Education Science. In BOSE Conf* (1–10).

Lindell, R. S., E. Peak, and T. M. Foster. 2007. Are They All Created Equal? A Comparison of Different Concept Inventory Development Methodologies. In AIP Conference Proceedings (Vol. 883, No. 1, pp. 14–17). American Institute of Physics.

Mahaffy, Peter G., Thomas A. Holme, Leah Martin-Visscher, Brian E. Martin, Ashley Versprille, Mary Kirchhoff, Lallie McKenzie, and Marcy Towns. 2017. "Beyond "Inert" Ideas to Teaching General Chemistry from Rich Contexts: Visualizing the Chemistry of Climate Change (VC3)." *Journal of Chemical Education* 94 (8): 1027–1035. https://doi.org/10.1021/acs.jchemed.6b01009

McCune, B., and M. J. Mefford. 2018. *PC-ORD. Multivariate Analysis of Ecological Data. Version 7.09*. Corvallis, OR, US: Wild Blueberry Media, LLC.

McMichael, A. J. 2013. "Globalization, Climate Change, and Human Health." *The New England Journal of Medicine* 368 (14): 1335–1343. https://doi.org/10.1056/NEJMra1109341

McNeal, K. S., S. L. Walker, and D. Rutherford. 2014. "Assessment of 6-to 20-Grade Educators' Climate Knowledge and Perceptions: Results from the Climate Stewardship Survey." *Journal of Geoscience Education* 62 (4): 645–654. https://doi.org/10.5408/13-098.1

Momsen, J., E. B. Speth, S. Wyse, and T. Long. 2022. "Using Systems and Systems Thinking to Unify Biology Education." *CBE Life Sciences Education* 21 (2): es3. https://doi.org/10.1187/cbe.21-05-0118

Nehm, Ross H., Elizabeth P, BeggrowJohn E. Opfer, and Minsu Ha. 2012. "Reasoning about natural selection: diagnosing contextual competency using the ACORNS instrument." *The American Biology Teacher* 74 (2): 92–98. https://doi.org/10.1525/abt.2012.74.2.6.

Parratt, Steven R., Benjamin S. Walsh, Soeren Metelmann, Nicola White, Andri Manser, Amanda J. Bretman, Ary A. Hoffmann, Rhonda R. Snook, and Tom A. R. Price. 2021. "Temperatures That Sterilize Males Better Match Global Species Distributions than Lethal Temperatures." *Nature Climate Change* 11 (6): 481–484. https://doi.org/10.1038/s41558-021-01047-0

Patz, J. A., D. Campbell-Lendrum, T. Holloway, and J. A. Foley. 2005. "Impact of Regional Climate Change on Human Health." *Nature* 438 (7066): 310–317. https://doi.org/10.1038/nature04188

Reis, J., and R. C. Ballinger. 2020. "Creating a Climate for Learning-Experiences of Educating Existing and Future Decision-Makers about Climate Change." *Marine Policy* 111. https://doi.org/10.1016/j.marpol.2018.07.007

Rittel, H. W. J., and M. M. Webber. 1974. "Dilemmas in the General Theory of Planning." *Policy Sciences* 4 (2): 155–169. https://doi.org/10.1007/BF01405730

Salimi, S., S. A. Almuktar, and M. Scholz. 2021. "Impact of Climate Change on Wetland Ecosystems: A Critical Review of Experimental Wetlands." *Journal of Environmental Management* 286: 112160. https://doi.org/10.1016/j.jenvman.2021.112160

Srinivasan, U., and D. S. Wilcove. 2021. "Interactive Impacts of Climate Change and Land-Use Change on the Demography of Montane Birds." *Ecology* 102 (1): e03223. https://doi.org/10.1002/ecy.3223

Uminski, C., and B. A. Couch. 2021. "GenBio-MAPS as a Case Study to Understand and Address the Effects of Test-Taking Motivation in Low-Stakes Program Assessments." *CBE Life Sciences Education* 20 (2): ar20. https://doi.org/10.1187/cbe.20-10-0243

Walther, Gian-Reto, Eric Post, Peter Convey, Annette Menzel, Camille Parmesan, Trevor J. C. Beebee, Jean-Marc Fromentin, Ove Hoegh-Guldberg, and Franz Bairlein. 2002. "Ecological Responses to Recent Climate Change." *Nature* 416 (6879): 389–395. https://doi.org/10.1038/416389a

Wasendorf, Chloe, Joshua W. Reid, Rebecca Seipelt-Thiemann, Z. T. Grimes, Brock Couch, Nick T. Peters, Julia Massimelli Sewall, Audrey L. McCombs, Patrick I. Armstrong, and Nancy Boury. 2022. "The Development and Validation of the Mutation Criterion Referenced Assessment (MuCRA)." *Journal of Biological Education:* 1–15. https://doi.org/10.1080/00219266.2022.2100451

Wilson, Christopher D., Charles W. Anderson, Merle Heidemann, John E. Merrill, Brett W. Merritt, Gail Richmond, Duncan F. Sibley, and Joyce M. Parker. 2006. "Assessing Students' Ability to Trace Matter in Dynamic Systems in Cell Biology." *CBE Life Sciences Education* 5 (4): 323–331. https://doi.org/10.1187/cbe.06-02-0142

Wise, S. B. 2010. "Climate Change in the Classroom: Patterns, Motivations, and Barriers to Instruction among Colorado Science Teachers." *Journal of Geoscience Education* 58 (5): 297–309. https://doi.org/10.5408/1.3559695