Concept Inventory Design for Determining Students’ Conceptual Understanding of Oceanography

Leilani Arthurs

University of Nebraska-Lincoln, larthurs2@unl.edu

Follow this and additional works at: https://digitalcommons.unl.edu/dberspeakers

Part of the Curriculum and Instruction Commons, Educational Methods Commons, Higher Education Commons, and the Science and Mathematics Education Commons

Arthurs, Leilani, "Concept Inventory Design for Determining Students’ Conceptual Understanding of Oceanography" (2012). DBER Speaker Series. 7.
https://digitalcommons.unl.edu/dberspeakers/7

This Article is brought to you for free and open access by the Discipline-Based Education Research Group at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in DBER Speaker Series by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Abstract for DBER Group Discussion on 2012-01-17

Presenter, Department(s):
Leilani Arthurs
Department of Earth & Atmospheric Sciences
Affiliation: Center for Science, Mathematics, & Computer Education

Title:
Concept Inventory Design for Determining Students’ Conceptual Understanding of Oceanography

Abstract:
Concept inventories are relatively new types of diagnostic instruments intended to measure student learning. Concept inventories exist for astronomy, biology, chemistry, engineering, fluid mechanics, geology, and physics. None is yet available for oceanography, and development of the Introductory Oceanography Concept Inventory Survey (IO-CIS) serves to fill this gap. A context-driven development strategy was designed to create this instrument. Qualitative methods utilizing grounded theory and classical test theory were used to construct it. Quantitative methods, including statistical methods associated with classical test theory and item response theory, were used to evaluate and further refine the IO-CIS. The instrument is valid and reliable for the population for which it was designed, a large west-central introductory oceanography course, and has potential to be used with and expanded upon for broader populations. In addition, the strategy and methods used to
Concept Inventory Design for Determining Students’ Conceptual Understanding of Oceanography
As scientists and science educators, we are very familiar the idea and practice of measuring things.

- How do we measure what students think and are able to do?
- Tests = measures of “constructs”
  - A construct is “the concept or characteristic that a test is designed to measure”¹.
- Examples²
  - The understanding of astronomical concepts
  - The ability to design a scientific experiment

¹ American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME), 1999
² Briggs, C.B., Alonzo, A.C., Schwab, C., and Wilson, M., 2006
Various methods at different times during a semester can be used.

- **Formative Assessments**
  - In-class exercises
  - Clicker questions
  - Lecture tutorials
  - Weekly quizzes or check-in reports

- **Summative Assessments**
  - Final projects or final reports
  - Exams
  - Concept inventories
What is a concept inventory?

- Diagnostic instrument
- Multiple-choice test
- Incorporates novice ideas and notions

Question or statement goes here (i.e., the stem).

A. Distractor 1
B. Distractor 2
C. Distractor 3
D. Distractor 4
E. Correct or most expert-like answer

- Pre/Post; formative/summative; learning gains
Concept inventories are a relatively new type of assessment instrument.

<table>
<thead>
<tr>
<th>Year</th>
<th>CI Construct</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>Electricity &amp; magnetism</td>
<td>Maloney, et al.</td>
</tr>
<tr>
<td>2005</td>
<td>Geoscience</td>
<td>Libarkin &amp; Anderson</td>
</tr>
<tr>
<td>2006</td>
<td>Greenhouse effect</td>
<td>Keller</td>
</tr>
<tr>
<td>2008</td>
<td>Genetics</td>
<td>Smith, et al.</td>
</tr>
</tbody>
</table>

This is not an exhaustive list of available concept inventories (CIs). The 1992 CI was the first to be developed and many others have been developed since that time.
How are concept inventories (CIs) developed?

- No single agreed upon method

- Usually based on research into students’ common reasoning difficulties

[Grounded theory used as theoretical framework to research student reasoning difficulties.]
Develop a concept inventory (CI) for oceanography.

- **Goal**
  - Use it to gauge student understanding of oceanography before and after instruction
  - Develop for course at large west central university

- **Stages of test construction**
  - Development stages: Qualitative methods
  - Evaluation stages: Quantitative methods
Earliest stages defined what the concept inventory (CI) would test for.

[These steps simplify the iterative method we developed and that is detailed in the handout.]

- **Step 1:** Instructor provided list of concepts he wanted students to know about (\textit{vocab list})
  
- **Step 2:** Learning objectives were written for each one of the listed concepts (\textit{what able to do})
  
- **Step 3:** Developed open-ended questions aligned with the learning objectives
Subsequent stages addressed validation and reliability of questions.

- Step 4: Student & expert interviews to “vet” Qs
- Step 5: Administer in-class exercise with Qs
- Step 6: Sort student responses by category
- Step 7: Discard or keep open-ended Qs
Subsequent stages addressed validation and reliability of questions.

- Step 8: Use student responses to develop answers
- Step 9: Convert open-ended Qs into multiple-choice Qs
- Step 10: Conduct student and expert interviews
- Step 11: Refine Q&A language and figures → 23 multiple-choice Qs
Phases and Methods Involved in the Construction of the Introduction to Oceanography Concept Inventory Survey (IO-CIS)

<table>
<thead>
<tr>
<th>Development Phases</th>
<th>Evaluation Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phase I.</strong></td>
<td><strong>Phase II.</strong></td>
</tr>
<tr>
<td>2. Define learning goals</td>
<td>11. Refine Q&amp;A language and figures, and interview experts and novices with modified Q&amp;As</td>
</tr>
<tr>
<td>3. Develop &amp; refine CIE* open-response Qs</td>
<td></td>
</tr>
<tr>
<td>4. Interview experts and novices with CIE* Qs</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. Administer CIE*</td>
</tr>
<tr>
<td></td>
<td>6. Code &amp; bin CIE*</td>
</tr>
<tr>
<td></td>
<td>7. Discard or keep CIE* questions</td>
</tr>
<tr>
<td>Cycle for each course module (i.e. 1 CIE/module)</td>
<td>8. Use student language to write answer options</td>
</tr>
<tr>
<td></td>
<td>9. Convert CIE* Qs into multiple-choice Qs, and interviews experts and novices with converted Qs</td>
</tr>
</tbody>
</table>

*CIE: Concept Inventory Exercise
### After development, the CI was used in three different sections (total n=464)

- Statistical analyses of student responses were used to evaluate individual Qs and overall CI

<table>
<thead>
<tr>
<th>Section</th>
<th>Instructor</th>
<th>Pre/Post</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>Post</td>
<td>81</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>Pre</td>
<td>152</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>Post</td>
<td>139</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>Pre</td>
<td>92</td>
</tr>
</tbody>
</table>
Classical Test Theory (CTT) has been traditionally used to evaluate CIs.

- **Assumptions:**
  1. Unidimensionality
  2. Observed score = Actual score + Error

- **Statistics:**
  1. Internal consistency & reliability (Kuder-Richardson Formula 20; Cronbach alpha)
  2. Item difficulty
  3. Item percent discrimination
  4. Instrument discrimination (COES)
What did CTT statistics tell us about the oceanography CI?

- Internally consistent & reliable
  - KRF20 = 0.57
  - Cronbach alpha = 0.73

- Discriminating & good length
  - COES = 0.51 (goal btw. 0.3 and 0.6)
  - Spread of item difficulty (70% spread)
  - Spread of item percent discrimination (50% spread)
Spread in Item Difficulty
[Example of what the data looks like]
Spread in Item Discrimination

[Example of the data]
Item Response Theory (IRT) is rarely used to evaluate CIs.

- Assumptions:
  - (1) Unidimensionality
  - (2) Local independence

- Statistical models:
  - (1) 1-parameter (Rasch)
  - (2) 2-parameter
  - (3) 3-parameter
Item Characteristic Curve [Examples of the data]
What were the results of IRT analysis (2-parameter model)?

- Shorter CI: 7 Qs out
  - 1 deemed too difficult
  - 6 have interchangeable difficulty levels with other Qs

- Can create different 16-item versions of the CI by exchanging Qs with similar difficulty levels
What did IRT statistics (2-parameter model) tell us about the CI?

- Internally consistent (Cronbach alpha = 0.69)
- 16-item CI provided a slightly better fit to the IRT model than the 23-item CI
- Spread of item difficulty (3.59 normal deviates)
- Spread of item discrimination (0.92)
- Can be further refined with greater response pool
## CTT vs. IRT

<table>
<thead>
<tr>
<th>Potential Pluses</th>
<th>Potential Minuses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CTT</strong></td>
<td><strong>Ability &amp; difficulty</strong></td>
</tr>
<tr>
<td>Simple stats</td>
<td>Sample dependent (recalc)</td>
</tr>
<tr>
<td>Detect poor Qs with even low n</td>
<td></td>
</tr>
<tr>
<td><strong>IRT</strong></td>
<td><strong>Larger n</strong></td>
</tr>
<tr>
<td>Ability &amp; difficulty</td>
<td>Sample independent</td>
</tr>
<tr>
<td>Sample independent</td>
<td>Involved stats</td>
</tr>
</tbody>
</table>
Conclusions

- CTT & IRT statistics both used to evaluate IO-CIS
- Statistical evidence for reliability
- Validity argument based on qualitative methods used in development phases
- Has potential to be refined for wider use
- Overall approach developed for constructing the IO-CIS is transferable to the development of concept inventories in other areas and disciplines.
Acknowledgements

- Science Education Initiative for funding
- CU-B, GSU, & UNL for support of the research
- All participating students
- Thomas Marchitto, CU-Boulder
- Baylor Fox-Kemper, CU-Boulder
- Derek Briggs, CU-Boulder
- Jennifer Hsia, University of South Dakota
- William Schweinle, University of South Dakota
Questions & Discussion ...

Possible Points of Discussion

(1) What is the value in rigorously developing assessments?
(2) To what extent do faculty admin pre/post assessments?
(3) What role does assessment in play under the umbrella of UNL’s ACE program, state, and national calls to improve STEM education?