Validation of an instrument measuring student complex causal assumptions

M. Shane Tutwiler, Tina Grotzer, Meredith Thompson, Amy Kamarainen, Shari Metcalf, Chris Dede

Harvard Graduate School of Education

Abstract

In this study, we present the validation evidence of a survey designed to assess student complex causal assumptions. We found that student responses could be explained by a two level measurement model, with responses to the complex causal domains of action at a distance, change over time, and non-obvious causes loading on a higher order factor, as well as a construct of agent-based causality. Reliability measures of the assessment when given before the instrument were adequate (alpha=0.71), and students showed expected gains in the measures after using a multi-user virtual environment (d=0.47, p<.001).

Problem

Students often fail to focus on deep patterns in ecosystems, instead focusing on surface-level structures (Hmelo-Silver, 2007). Recognizing this problem, researchers developed an ecosystems science focused multi-user virtual environment (MUVE) (Metcalf et al., 2011). In the MUVE, students work in teams of four to solve problems in one of two modules: a pond (fish die-off) and a forest (predator/prey balance). Assessment of student understanding of the types of causal patterns found in the MUVE was conducted via responses to open-ended items, coded iteratively to surface important trends (Grotzer et al., 2013).

In an effort to stream-line the assessment for future use with the latest version of the MUVE (Thompson et al., 2016), the authors have adopted the original open-ended survey into a set of sixteen five point Likert-type items called the Causal Assessment of Understanding in Systems & Ecosystems (CAUSE), meant to measure how strongly students believe it is important to collect data in support of four types of causal patterns: action at a distance vs. actions nearby, non-obvious vs. obvious causes, change over time vs. immediate change, and non-agent vs. agent-based. In this study, we explore the initial validity evidence (e.g., Ketelhut, 2010) from pilot testing of this instrument in the spring of 2015, fall of 2015, and spring of 2016, and highlight future work.

Procedure

In the spring of 2015 we pilot tested the survey with middle school students in a New England school district. The assessment was initially taken by 156 students, prior to their use of the MUVE. Students then used the MUVE curriculum for approximately 10 days. Finally, 144 of those students were present to take the assessment again after using the MUVE pond module. We tested the instrument with an additional 178 students in the fall of 2015 and 79 students in the spring of 2017.

Analyses & Findings

Spring 2015

To establish initial evidence of construct validity, we first confirmed the face and content validity of the questions used in the survey. The survey was based on a previously used open-ended assessment designed by a domain expert in complex causal understanding, and used to make valid inferences in previously published peer-reviewed research (Grotzer et al., 2013). The re-designed questions on the survey under study were iteratively reviewed by the same domain expert as well as other members of the research team familiar with k-12 science education and assessment. Examples of questions used in the survey are given in Figure 1. For example, students demonstrate more complex causal assumptions when they disagree with the statement "Looking for clues far away from the edge of the pond is a waste of time." Conversely, students who agree with this statement use less complex causal assumptions.

	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree
It is important to look for clues at the edge of the pond.					
It's a bad idea to only focus on the things we can see.					
It is important to search for close far away from the edge of the pond.					
Looking for clues far away from the edge of the pond is a waste of time.					
We need to focus on the day that we found the dead frogs.					
It's a bad idea to only look for clues from the day we found the dead frogs.					

Figure 1. Example questions from the survey

We next explored patterns of student responses to the items via principal components analysis. We hypothesized that students reflecting on multiple types of complex causal patterns would result in multiple strong first principal components (scree-plot values greater than 1). Note in Figure 2 that seven of the principal components derived from the full survey resulted in scree values greater than 1. A follow-up item analysis showed that six items had deleted-item alpha scores higher than the global alpha score suggesting that the reliability of the assessment was negatively impacted by their inclusion. These items (1, 2, 5, 6, 9, & 13) were removed for the remainder of the analyses and will be re-written and re-analyzed in the future. The overall reliability of this instrument, as measured by Cronbach's alpha, was 0.61 (Table 1.)

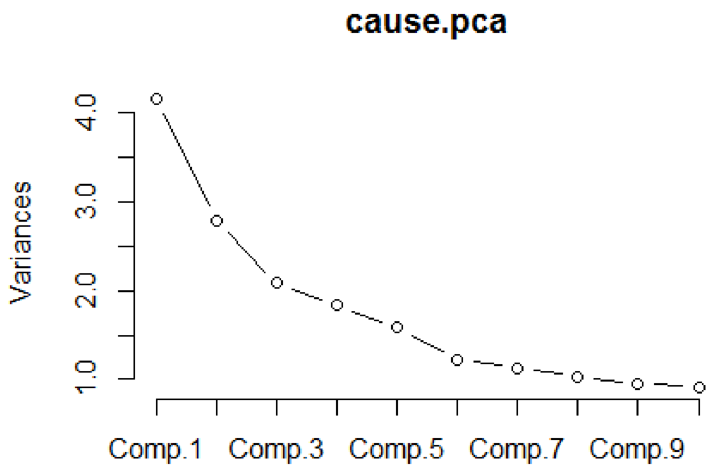

Figure 2. Principal Components Analysis of the pre-assessment (n=156)

A confirmatory factor analysis on the remaining reliable items resulted in adequate model fit (χ^2(38)=50.543, p=.084; CFA=0.933; RMSEA=.046(0.000-0.077), p=0.551; SRMR=.056) with multiple factor loadings near or above 1.0. We note in Figure 3 that a single, higher order factor (named "complex causal understanding" here, but yet to be independently validated) loaded strongly on the distance, temporal, and obviousness factors, but not the agency factor. This is in line with prior work (Grotzer & Tutwiler, 2014), which posited that a strong focus on agentive causes acts counter to other forms of complex causal assumptions.

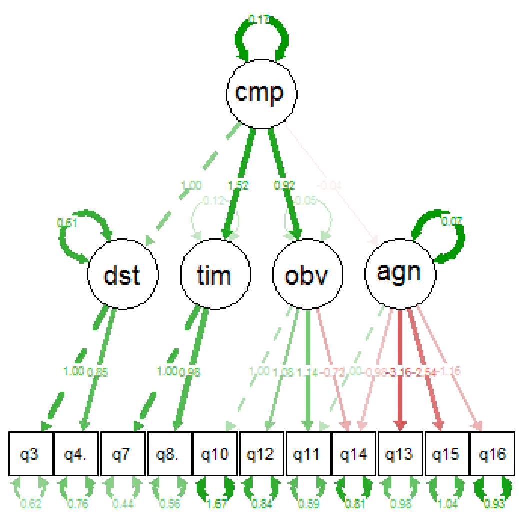

Figure 3. Confirmatory Factor Analysis of the reduced pre-assessment (n=156)

We also analyzed shifts in average scores on the instrument for a subset of 144 students who took both the pre and post assessments. We note first that the reliability score of the instrument shifted positively from 0.60 on the pre-assessment to 0.70 on the post-assessment.

Also, there was a moderate positive correlation (r=0.44) between the pre and post measures (Figure 4). Finally, we notice that average scores rose by 0.21 points, a statistically significant (t(143)=5.69 , p<.001) difference representing an effect size of 0.47 standard deviation units (Figure 5). This was expected, given the curricular nature of the MUVE, and gives further evidence that the construct measured by the survey is valid.

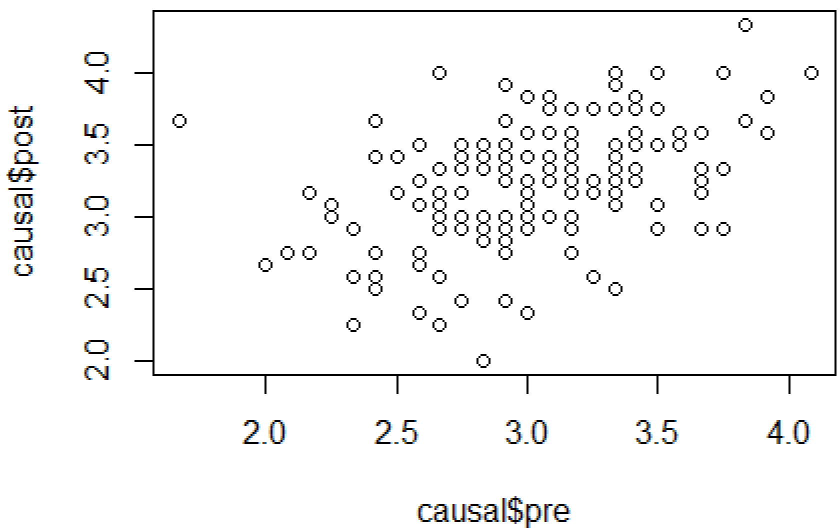

Figure 4. Scatter plot of pre and post intervention reduced survey scores (n=144)

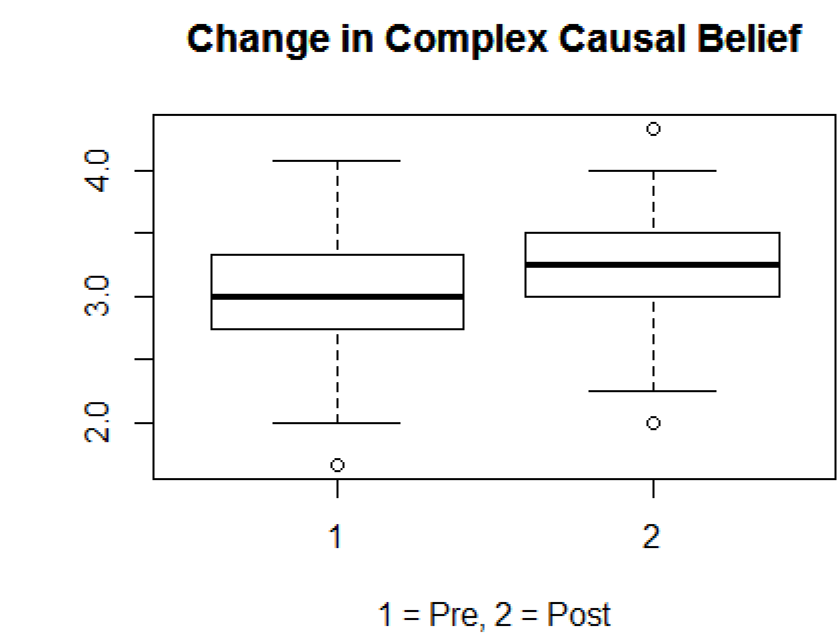

Figure 5. Average pre (m=3.03(0.43)) and post (m=3.24(0.43)) reduced survey scores (n=144)

Fall 2015

We edited questions that demonstrated low reliability (outlined above) and re-administered the survey to 178 middle school science students in the fall of 2015, specifically to assess overall reliability and construct validity (via CFA). As in the spring, a two-level factor model best fit the data (Figure 6.) Reliability of the instrument, via Cronbach's alpha, was 0.70 (Table1.) This model demonstrated adequate fit ($\chi 2(38)$=102.632, p=.303; CFA=0.982; RMSEA=0.20(0.000-0.045), p=0.979; SRMR=0.063). Interestingly, the agentive factor in this replication was not negatively related to the complex causal factor, as it was in the spring 2015 study. Further exploration of the relationship between these two factors will be conducted in future studies.

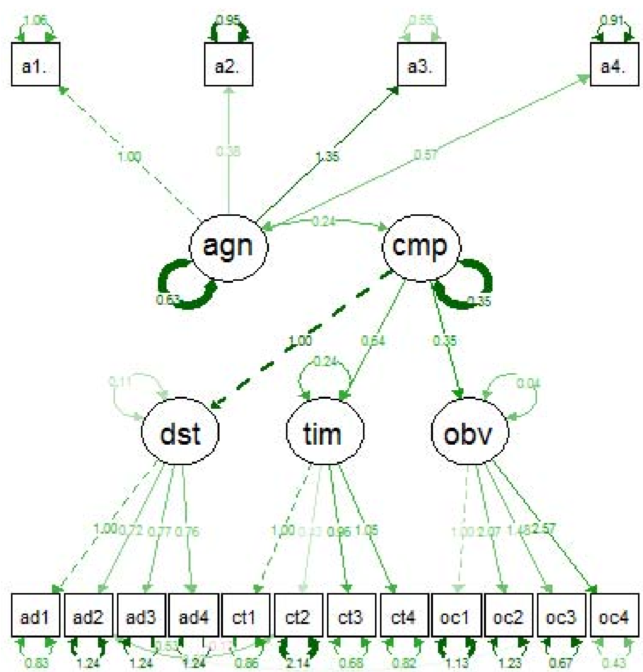

Figure 6. Confirmatory Factor Analysis of the revised pre-assessment (N=178)

Spring 2016

Analysis of the data from the Spring 2016 pilot test of the CAUSE showed adequate reliability (Table 1), though the assessment was given after the MUVE intervention, not before (Cronbach's alpha = 0.71).

Table 1. Reliability measures of pre-assessments across three validation studies of the CAUSE

Study 1	Study 2	Study 3
0.61	0.70	0.71

In summary, we have established preliminary evidence of the construct validity of the CAUSE survey. Future work, which will be conducted in fall 2016, will include correlation of these measures to external assessments in an effort to establish convergent and discriminant validity evidence.

Contributions

Once fully developed and validated, the survey will allow us to measure and assess changes in student complex causal assumptions. These insights will help us to understand how other students using similar curricula (supported by technology or otherwise) might experience shifts in their own beliefs about the importance of exploring various types of complex causal patterns. This is central to the process of science teaching and learning, as understanding of complex causality/complex systems is important at all levels of modern science standards such as the NGSS (NGSS Lead States, 2013).

References

Grotzer, T.A., Kamarainen, A., Tutwiler, M.S, Metcalf, S, & Dede, C. (2013) Learning to reason about ecosystems dynamics over time: The challenges of an event-based causal focus. *BioScience*. 63(4):288-296.

Grotzer, T.A., & Tutwiler, M.S. (2014) Simplifying Causal Complexity: How Interactions Between Modes of Causal Induction and Information Availability Lead to Heuristic-Driven Reasoning. *Mind, Brain, and Education*. 8(3):97-114.

Hmelo-Silver, C.E., Marathe, S., Liu, L. (20017). Fish swim, rocks sit, and lungs breathe: Expert-novice understanding of complex systems. *Journal of the Learning Sciences*. 16: 307-331.

Ketelhut, D.J. (2010). Assessing gaming, computer and scientific inquiry self-efficacy in a virtual environment. In L.A. Annetta and S. Bronack (Eds.), *Serious Educational Game Assessment: Practical Methods and Models for Educational Games, Simulations and Virtual Worlds.* Amsterdam, The Netherlands. Sense Publishers. p. 1-18.

Metcalf, S., Kamarainen, A., Tutwiler, M.S., Grotzer, T., & Dede, C. (2011). Ecosystem science learning via multi-user virtual environments. *International Journal of Gaming and Computer-Mediated Simulations*, 3(1) 86-90.

NGSS Lead States (2013). *Next Generation Science Standards: For States, By States.* Washington, DC: The National Academies Press.

Thompson, M.M., Tutwiler, M.S., Metcalf, S.J., Kamarainen, A.M., Grotzer, T.A, Dede, C.J. (2016) A blended assessment strategy for EcoXPT: An experimentation-driven ecosystems science based multi user virtual environment. Presented at the annual meeting for the American Education Research Association (AERA), Washington, DC. April, 2016