International Society of the Learning Sciences

# Understanding of a Law of Science and Its Relation to Science Writing with Automated Feedback

ChanMin Kim, The Pennsylvania State University, cmk604@psu.edu
Sadhana Puntambekar, The University of Wisconsin-Madison, puntambekar@education.wisc.edu
Eunseo Lee, The Pennsylvania State University, eul374@psu.edu
Dana Gnesdilow, The University of Wisconsin-Madison, gnesdilow@wisc.edu
Indrani Dey, The University of Wisconsin-Madison, srbaker2@wisc.edu
Xuesong Cang, The University of Wisconsin-Madison, xcang@wisc.edu
Siyu Wu, The Pennsylvania State University, sfw5621@psu.edu
Rebecca J. Passonneau, The Pennsylvania State University, rjp49@psu.edu

**Abstract:** Building causal knowledge is critical to science learning and scientific explanations that require one to understand the how and why of a phenomenon. In the present study, we focused on writing about the how and why of a phenomenon. We used natural language processing (NLP) to provide automated feedback on middle school students' writing about an underlying principle (the law of conservation of energy) and its related concepts. We report the role of understanding the underlying principle in writing based on NLP-generated feedback.

## Introduction

Understanding cause-and-effect relationships is an essential part of reasoning about scientific phenomena and writing scientific explanations. At the core of this process is learning underlying principles that explain observed/identified relationships (Russ et al., 2008; Vieira et al., 2019). This contributes to "a clear conceptual understanding of the principles and theories, plus the knowledge of how to apply these principles to a different context" (Vieira et al., 2019, p. 203). For example, when experimenting with a roller coaster simulation, students may be able to identify that a car's motion will be sustained through the remainder of the ride when the initial drop height is higher than a subsequent hill. However, they may not be able to explain why this is so, particularly when they do not understand the underlying principle.

In this study, middle school students learned about the law of conservation of energy (LCE) and its associated concepts and relationships (e.g., potential energy, kinetic energy, total energy), and received automated feedback through natural language processing (NLP) on their writing about the principle and related concepts. Our research questions were:

1. How does students' early understanding of an underlying principle (law of conservation of energy) relate to their writing about the principle and related concepts?
2. How does students' understanding of the underlying principle relate to feedback effectiveness (evidenced in the quality of the revised essay)?

## Conceptual framework

We constructed the conceptual framework of the present study based on literature recognizing intuitive knowledge (diSessa, 1988, 2018), intuitive theories (Gopnik, 2012), and intuitive explanations (Keil & Wilson, 2000) as part of a pathway toward scientific theory and scientific explanations. We acknowledge the importance of mechanistic reasoning (Carmichael et al., 2010; Russ et al., 2008; Vieira et al., 2019) in scientific explanations in that it enables one to understand underlying principles. Given that scientific explanations describe the how and why of a phenomenon based on scientific facts (Osborne & Patterson, 2011), covariational reasoning, which identifies cause-and-effect relationships without the why, is not sufficient. At the same time, we value the learning process that includes covariational reasoning and even perceptual explanations, especially considering that causal knowledge involves both scientific and intuitive theories (Gopnik, 2012). The conceptual framework guides us to attend more to the process of "reconciling theory with experience of the natural world" (Furtak et al., 2010, p. 177). We conceptualize automated feedback from NLP in the present study as a scaffold for students to translate between underlying principles and experience during the reconciliation process.

## Method

Two eighth-grade science teachers and their 138 students from a mid-sized, US Midwestern city participated in this study. The study was conducted during a three-week, design-based physics unit focused on energy conservation and transformation, where students were invited to design a roller coaster using what they learned about physics. During the unit, students participated in five virtual labs using a roller coaster simulation, answered questions after each lab, and wrote and revised two design essays. The initial design essay (Essay 1) was to explain how height and mass affected the amount of energy there was in their roller coaster system, as well as how the LCE could be used to explain transformations. This initial essay was sent to our NLP technology, called PyrEval, to automatically assess students' essays for feedback, which students could later use to revise their ideas. Students then learned about how height and mass affected speed. Students then wrote their second design essay (Essay 2), building on the ideas in their initial essay, using the feedback from PyrEval. Automated feedback was generated through the NLP technology that detected four main content units in students' essays: one was about LCE and the other three were about related concepts (i.e., height and energy; mass and energy; initial drop in relation to hill height). Based on the presence or absence of each content unit, the feedback was given to either acknowledge their inclusion of the content unit or prompt them to explain the missing content unit. For example, students who did not explain LCE but explained other related concepts in their Essay 1 were asked to elaborate on their current explanation in connection with LCE.

Data for this study consisted of students' written responses from short answer questions about LCE after labs and for their design essays. Two researchers independently coded 15% of all students' responses and achieved an Intra Class Correlation value of .947, which is considered excellent (Cicchetti, 1994). All discrepancies between raters were resolved through discussion.

## Findings and discussion

### RQ1: How does students' early understanding of the underlying principle (LCE) relate to their writing about the principle and the related concepts?

We sorted students into four groups depending on which of the early short answer questions about LCE they answered correctly. We also categorized Essay 1 into four explanation patterns depending on which of LCE and related concepts were explained. We then ran a Fisher's exact test. Table 1 lists all groups and patterns.

**Table 1**
*Early Understanding Groups and Essay 1 Explanation Patterns*

| **Early Understanding of the Underlying Principle:** Answers to the first and second lab questions about LCE | | **Essay 1 Explanation Patterns:** Explanations of LCE and related concepts | |
|---|---|---|---|
| **Group A** (n=29) | Correctly answered the first and second lab questions | **Pattern 1** (n= 70) | Wrote about LCE and at least one of the related concepts |
| **Group B** (n=23) | Correctly answered only the first lab question | **Pattern 2** (n= 4) | Wrote about LCE |
| **Group C** (n=32) | Correctly answered only the second lab question | **Pattern 3** (n=51) | Wrote about at least one of the related concepts |
| **Group D** (n=54) | Did not answer either question correctly | **Pattern 4** (n=13) | Wrote neither |

Fisher's exact test results showed that there was a statistically significant association between students' early understanding of LCE and their Essay 1 writing quality (two-tailed, $p<.001$). Observed frequency table (Table 2) shows that among students who answered both lab questions about LCE correctly (Group A), 65.5% explained LCE *and* at least one related concept correctly in Essay 1. Among students who answered one of the lab questions about LCE correctly (Groups B and C), 65.6%-73.9% explained LCE *and* at least one of the related concepts correctly in their Essay 1. Among students who were not able to answer any of the lab questions about LCE correctly (Group D), only 24.1% explained LCE *and* at least one of the related concepts correctly in Essay 1, but 64.8% explained at least one of the related concepts correctly in their Essay 1. The results also show that it was possible to write about related concepts without understanding the underlying principle. For example, one of the related concepts that students were expected to write about in their essays was the initial drop height of the roller coaster being higher than the subsequential hill that they designed to get the car to travel to the end of the roller coaster. That is, without understanding of LCE, cause-and-effect relationships were discoverable on the simulation and the data summary table without knowledge of underlying principles. Such phenomenological explanations (Furtak et al., 2010) written by students in Group D are not deficits within our conceptual framework. They are the opportunity for the process of reconciling underlying principles with data to begin. These students received automated feedback on their Essay 1 asking them to write about LCE.

**Table 2**

*Observed Frequencies of Essay 1 Explanation Patterns per Early Understanding Group*

| | **Pattern 1** LCE and related concepts | **Pattern 2** LCE only | **Pattern 3** Related concepts only | **Pattern 4** No LCE and no related concepts |
|---|---|---|---|---|
| **Group A** | 65.5% (n=19) | 0.0% (n=0) | 24.1% (n=7) | 10.3% (n=3) |
| **Group B** | 73.9% (n=17) | 4.3% (n=1) | 8.7% (n=2) | 13.0% (n=3) |
| **Group C** | 65.6% (n=21) | 9.4% (n=3) | 21.9% (n=7) | 3.1% (n=1) |
| **Group D** | 24.1% (n=13) | 0.0% (n=0) | 64.8% (n=35) | 11.1% (n=6) |

*Note*. Output in each cell indicates percentage within each early understanding group.

## RQ2: How does students' understanding of the underlying principle (LCE) relate to feedback effectiveness (evidenced in the quality of the revised essay)?

We categorized Essay 2 into the four explanation patterns as in the first Fisher's exact test and used the same early understanding group data. We then ran another Fisher's exact test and found a statistically significant association between students' early understanding of LCE and their Essay 2 writing quality (two tailed, $p<.001$) (Table 3).

**Table 3**

*Observed Frequencies of Essay 2 Explanation Patterns per Early Understanding Group*

| | **Pattern 1** LCE and related concepts | **Pattern 2** LCE only | **Pattern 3** Related concepts only | **Pattern 4** No LCE and no related concepts |
|---|---|---|---|---|
| **Group A** | 72.4% (n=21) | 0.0% (n=0) | 20.7% (n=6) | 6.9% (n=2) |
| **Group B** | 65.2% (n=15) | 0.0% (n=0) | 21.7% (n=5) | 13.0% (n=3) |
| **Group C** | 81.3% (n=26) | 6.3% (n=2) | 12.5% (n=4) | 0.0% (n=0) |
| **Group D** | 37.0% (n=20) | 0.0% (n=0) | 57.4% (n=31) | 5.6% (n=3) |

*Note*. Output in each cell indicates percentage within each early understanding group.

Overall improvements in writing quality were visible in that more essays included correct explanations of both LCE and related concepts (Pattern 1), and less essays included no correct explanation of LCE and/or related concepts (Pattern 4) in Essay 2 than Essay 1. Automated feedback that prompted students to write about LCE when other related concepts were explained without explaining the why, beyond referring to their simulation data as reasons, may have helped students connect concrete experiences (from the simulation) to LCE. These findings also suggest that early understanding of LCE could be impactful, but based on Group C who showed the largest increase in Pattern 1, a longer process of reconciling theory with data may have been even better.

The overall improvement in explanation patterns hinted that the feedback given between Essay 1 and Essay 2 may have played a positive role. We ran a repeated measures ANOVA to see if there was a statistically significant difference in writing quality scores between Essay 1 and Essay 2. The results showed that Essay 2 was significantly better than Essay 1, $F(1, 134) = 22.96$, $p<.001$. The effect size was medium (Cohen's $d = .42$). Furthermore, the improvement from Essay 1 to Essay 2 differed depending on the explanation patterns that the students included in Essay 1, $F(3, 134) = 101.77$, $p<.001$, Cohen's $d = 1.5$. That is, the improvement from Essay 1 to Essay 2 shown among students who included explanations of only LCE (Pattern 2) in Essay 1 was significantly larger than the improvement shown among the students who included explanations of only related concepts (Pattern 3) in Essay 1. Feedback worked better among those who understood and wrote about the underlying principle (LCE) than those who were able to write about one or more of related concepts but without the underlying principle. The effect size was large (Cohen's $d = 1.5$).

We also ran linear mixed effect models using the lme4 R package (Bates et al., 2014) to see further about the relations between students' understanding of the underlying principle and their revised essay, with other possible predicting variables for the revised essay quality. Table 4 lists the output of mixed effect model analysis with Essay 2 writing quality as a dependent variable. We included fixed effects of early understanding group, Essay 1 explanation pattern, Essay 1 writing quality, NLP accuracy, Essay 1 revision, and engagement. We also included in the model teacher and class clustering factors as random effects to control for the potential impact of the teacher and class variance. The model specification was as follows: Essay 2 writing quality ~ early understanding group + Essay 1 explanation pattern + Essay 1 writing quality + Essay 1 revision + engagement + (1|Teacher) + (1|Class). There were two significant predicting variables at .05 significance level: Essay 1 writing quality ($ß=0.87$, $p<0.0001$) and Essay 1 revision ($ß=1.12$, $p<0.0001$). While there were still indirect effects of early understanding and Essay 1 explanation pattern reported above, only these two variables were direct predictors for Essay 2 writing quality. This means that students' revised essay quality was better when they revised their Essay 1 as per automated feedback. It seems intuitive that their revised essay quality was better when their first essay quality was already better, but this finding also suggests that the automated feedback did not ask students to revise their essay when unneeded. Also, the finding that their revised essay quality was better when revisions were made according to the automated feedback suggests that the positive impact of automated feedback on improving writing quality. Especially considering the improvements reported above including more Explanation Pattern 1 in Essay 2, the findings demonstrate a unique potential contribution of PyrEval to science learning and writing as a scaffold for students' translating, connecting, and reconciling between theory with experience (diSessa, 2018; Furtak et al., 2010; Puntambekar & Goldstein, 2007). This will in turn contribute to knowledge building that recognizes possible interplay between intuitive explanations and scientific explanations and value the role of intuitive explanations that can be leveraged through automated feedback scaffolding toward scientific explanations.

**Table 4**
*Linear Mixed Effects Model Analysis results for Essay 2 Writing Quality Scores*

| | Estimate | Std. Error | t value | Pr(>\|t\|) | |
|---|---|---|---|---|---|
| (Intercept) | 0.2564 | 0.37739 | 0.679 | 0.498 | |
| Early understanding group | -0.06933 | 0.0457 | -1.517 | 0.132 | |
| Essay 1 explanation pattern | 0.01206 | 0.08241 | 0.146 | 0.884 | |
| **Essay 1 writing quality** | 0.87115 | 0.07475 | 11.655 | 0.00002 | *** |
| NLP accuracy | 0.01633 | 0.05768 | 0.283 | 0.778 | |
| **Essay 1 revision** | 1.12133 | 0.11685 | 9.597 | 0.00002 | *** |
| Engagement | 0.10284 | 0.13072 | 0.787 | 0.433 | |
| Marginal $R^2$/Conditional $R^2$ | 0.691/0.803 | | | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## References


Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting linear mixed-effects models using lme4* (arXiv:1406.5823). arXiv. https://doi.org/10.48550/arXiv.1406.5823

Carmichael, A., Chini, J. J., Rebello, N. S., & Puntambekar, S. (2010). Comparing student learning in mechanics using simulations and hands-on activities. In C. Singh, M. Sabella, & S. Rebello (Eds.), *2010 Physics Education Research Conference Proceedings* (pp. 89–92).

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290.

diSessa, A. A. (1988). Knowledge in pieces. In *Constructivism in the computer age* (pp. 49–70). Lawrence Erlbaum Associates, Inc.

diSessa, A. A. (2018). A friendly introduction to "knowledge in pieces": Modeling types of knowledge and their roles in learning. In G. Kaiser, H. Forgasz, M. Graven, A. Kuzniak, E. Simmt, & B. Xu (Eds.), *Invited Lectures from the 13th International Congress on Mathematical Education* (pp. 65–84). Springer.

Furtak, E., Hardy, I., Beinbrech, C., Shavelson, R. J., & Shemwell, J. T. (2010). A framework for analyzing evidence-based reasoning in science classroom discourse. *Educational Assessment*, *15*(3–4), 175–196.

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science*, *337*(6102), 1623–1627.

Keil, F. C., & Wilson, R. A. (Eds.). (2000). *Explanation and cognition*. MIT Press.

Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education*, *95*(4), 627–638.

Puntambekar, S., & Goldstein, J. (2007). Comparing classroom enactments of an inquiry curriculum: Lessons learned from two teachers. *The Journal of the Learning Sciences*, *16*(1), 81–130.

Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, *92*(3), 499–525.

Vieira, C., Magana, A. J., Roy, A., & Falk, M. L. (2019). Student explanations in the context of computational science and engineering education. *Cognition and Instruction*, *37*(2), 201–231.


## Acknowledgement

This work was supported by grants 2010351 and 2010483 from the National Science Foundation (USA). Any opinions, findings, or conclusions are those of the authors and do not necessarily represent official positions of the National Science Foundation.