J. Korean Soc. Math. Ed. Ser. D. (2022) 25(2), 135–158 ISSN 1226-6191
https://doi.org/10.7468/jksmed.2022.25.2.135 Online ISSN 2287-9943

**RESEARCH ARTICLE**

# Building a Model(s) to Examine the Interdependency of Content Knowledge and Reasoning as Resources for Learning

Ali Cikmaz[1], Jihyun Hwang[2], Brian Hand[3]

[1] Adjunct Assistant Professor, Teaching and Learning, University of Iowa
[2] Assistant Professor, Mathematics Education, Korea National University of Education
[3] Professor, Teaching and Learning, University of Iowa

Received: June 7, 2022 / Accepted: June 29, 2022 / Published online: June 30, 2022


## Abstract

This study aimed to building models to understand the relationships between reasoning resources and content knowledge. We applied Support Vector Machine and linear models to the data including fifth graders' scores in the Cornel Critical Thinking Test and the Iowa Assessments, demographic information, and learning science approach (a student-centered approach to learning called the Science Writing Heuristic [SWH] or traditional). The SWH model showing the relationships between critical thinking domains and academic achievement at grade 5 was developed, and its validity was tested across different learning environments. We also evaluated the stability of the model by applying the SWH models to the data of the grade levels. The findings can help mathematics educators understand how critical thinking and achievement relate to each other. Furthermore, the findings suggested that reasoning in mathematics classrooms can promote performance on standardized tests.

**Keywords** Mathematics Reasoning, Science Heuristic Writing, Reasoning Resources, Interdependency between Reasoning and Content.

• This research was supported by a grant from the Iowa Measurement Research Foundation (IMRF)
• Corresponding Author: Jihyun Hwang, email: jihyun-hwang@knue.ac.kr

# I. INTRODUCTION

For better understanding of how individuals come to know science and mathematics, increasing attentions have been paid to students' practices (Yackel & Hanna, 2003) suggested in the national curriculum in mathematics and science – the Common Core State Standards for Mathematics (CCSSM; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) and the Next Generation Science Standards (NGSS; NGSS Lead States, 2013). These curricula have particularly emphasized reasoning as a way to generate various types of knowledge and the outcome of learning is content knowledge viewed as all the different forms of such knowledge. This indicates that reasoning necessarily entails use of epistemological practices, like using language, building arguments, and participating in dialogical interaction.

This perspective – reasoning generates content knowledge of disciplines – suggests that it is impossible to characterize students' reasoning without knowing what content knowledge they use. In other words, reasoning and content knowledge are interdependent. Interdependency between reasoning and content knowledge suggests that content knowledge is not a thing consumed in the process of reasoning or critical thinking. Rather, students utilize and develop the two types of intellectual resources in learning any discipline, content knowledge and reasoning resources (Sadler & Zeidler, 2005). When engaging in the epistemological practices above, learners' reasoning starts from inductive/abductive reasoning resources – select mathematical rules, make a hypothesis, and analyze patterns in given problem situations. Then, deductive reasoning resources were utilized next – generate a logical proof, representing a solution, and draw conclusions. The shift from inductive/abductive to deductive reasoning resources is critical in the process of developing content knowledge because students' understanding of any ideas should remain tentative with only inductive/abductive resources, but deductive resources make their knowledge explicit with justification.

The purpose of this research is to examine the relationships between the different types of intellectual resources – reasoning resources and content knowledge. Students' utilization of reasoning resources was evaluated via the Cornell Critical Thinking Test (CCTT; Ennis et al., 2004), while students' content knowledge was examined via the Iowa Assessments. We constructed statistical models representing these relationships, thus it would provide additional evidence about linearity and stability of the relationships between critical thinking and achievement, which have been examined by previous studies (e.g., Ganizadeh, 2017). Furthermore, the interest of this study is in individuals' utilization of resources and achievement. However, we also recognized that reasoning is normative and depends on learning environments. Thus, this research is expected to contribute to better understanding about the influence of learning environments on the relationships between reasoning and academic achievement.

# II. LITERATURE REVIEW

## Philosophical Perspective on Reasoning

Reasoning as practices for learning are one of physical and mental behaviors that experts generate their knowledge and theories (NGSS Lead States, 2013). Researchers have recognized that reasoning has a significant role in students' learning. Aligned with researchers' acknowledgement, in mathematics education the NCTM's (2000) principles and standards for school mathematics and the following CCSSM (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) documented what students should know and do in K-12 classrooms. For example, CCSSM suggested that students should have opportunities to reason abstractly and they are required to communicate precisely with their peers. The Next Generation Science Standards (NGSS Lead States, 2013) also recommended that students' opportunities to learn science should be shaped around content and practice standards together, for example, asking questions, planning and carrying out investigation, and draw conclusions.

Reasoning is necessarily involved in epistemological practices in the process of learning the content of the discipline because experts in mathematics or science reason to generate knowledge. However, mathematics and science have different epistemological frames: science is framed around causal explanation which is "contingent and demonstrated through evidence" (Moshman, 2015, p. 74), whereas mathematics is centered on "rule-based reasoning" and yields "objective truths" (p. 74). Advancing knowledge in mathematics and science is framed in different ways, but essentially both rely on using arguments, which is exploring the relationship between premise and conclusion (Blair, 2016; Woods, 2016). This relationship is framed by two different reasoning processes: searching or reasons and giving reasons (Kirwan, 1995).

The process of searching for reasons in science is abductive in nature (Lawson, 2005). Scientists need to make predictions with the best possible explanations to a phenomenon and design experiments to test their prediction. In mathematics, inductive reasoning is used to determine and select the rules (axioms, lemma, and existing theorems) to generate a logical proof connecting given conditions to the target concept or theorem. As Khomenko (2016) argued, "all the good arguments are being reduced to deductive ones" (p. 621), which indicates that there is a need of the process of giving reasons in order to ensure the initial conditions logically reach a conclusion.

While reasoning is considered as practices to generate knowledge in mathematics and science, other researchers have attempted to characterize and measure students' domain-general reasoning ability. One of such attempts is the development of the Cornell Critical Thinking Test (Ennis et al., 2004), which is based on the following: "Now if we set about to find out what … [a] statement means and to determine whether to accept or reject it, we would be engaged in thinking which, for lack of a better term, we shall call critical thinking" (Smith, 1953, p. 130). As the two different reasoning processes are characterized as inductive/abductive and deductive inferences, Ennis et al. (2004) suggests three types of inferences to beliefs (induction, deduction, and value judging) and four types of bases for those inferences (the results of other inferences, observations, statements made by others, and assumptions; pp. 1–2). This argument shows on what reasoning could base in addition to the types of reasoning.

## Relationship between Reasoning and Academic Achievement

As discussed in the introduction, the two types of the intellectual resources, content knowledge and reasoning resources, are interdependent in learning any discipline (Sadler & Zeidler, 2005). Content knowledge is not a resource consumed in the process of reasoning or critical thinking. Furthermore, it is impossible to characterize students' reasoning without knowing what content knowledge they use. Broadly, students' reasoning in mathematics usually relies on induction and deduction while science inquiry requires abduction and deduction. Even within a discipline like mathematics, proving the Pythagorean Theorem visually differs from justifying the quadratic formula by changing quadratic equations with some rules. On the other hand, students' content knowledge cannot be developed without reasoning (e.g., recalling a definition or representing). Thus, students' learning requires use of both content knowledge and reasoning resources interdependently.

Critical thinking requires content knowledge and reasoning resources in a given problematic situation, which "always takes place in response to a particular task, question, problematic situation or challenge" (Bailin, 2002, p. 368). This indicates that the interdependency between content knowledge and reasoning depends on learning environments. Particularly, students have different opportunities to use their own language to express their ideas, to engage in different realistic problems to construct their arguments, and participate in different degrees of dialogical interactions and collaborations to improve and criticize the arguments. Therefore, student-centered learning environments (the SWH in this research) provide rich opportunities to use content knowledge and reasoning resources with epistemic tools can change – confidently strengthen – the degree of the interdependency between content knowledge and reasoning. In addition, students' grade levels may influence the interdependency between reasoning and content knowledge. This is because different content knowledge is supposed to learn based on curriculum as well as different levels of complexity in students' reasoning are required (Hyman, 2015). Students reported that mathematics becomes very difficult in middle schools and academic declines due to transition from elementary to middle schools have been found (Elias, 2001, Winter; Theriot & Dupper, 2010).

The relationships between reasoning and academic achievement have not been documented, but critical thinking as a consequence of applying reasoning resources has been discussed in connection to academic achievement. Theoretically, Lunenburg (2011) argued that critical thinking and constructivist approaches are promising to improve academic achievement of all students in the core subject areas. Empirically, the positive relation between critical thinking and academic achievement seems quite obvious, but this relation has been directly compared in only a small number of studies (Stupnisky et al., 2008; Ghanizadeh, 2017). Ganizadeh (2017). These studies showed a high correlation between critical thinking and students GPA generally as a result of an implementation or intervention. Furthermore, these studies examined critical thinking and achievement separately, which means the relation between them was not examined in detail. Both critical thinking and achievement were improved regarding implementation (i.e., Yang & Chang, 2013; Azar, 2010).

Many empirical studies showed that critical thinking can be developed various student-centered interventions (Yang & Chang, 2013). Particularly about the SWH approach, there are a large body of literature showing the positive effects of this approach on academic achievement and critical thinking (Stephenson & Sadler-McKnight, 2016; Kingir et al., 2012; Akkus et al., 2007; Taylor et al., 2018; Hand et al., 2018; Tseng, 2014). These studies also showed improvement of academic achievement and critical thinking, but there is no direct link between achievement and critical thinking. This research gap raises question about how critical thinking and achievement are related in detail when students learning science via SWH environments.

The relationships between students' reasoning and achievement can depend on development of reasoning, but it is uncertain how the relationships differ by grade level. Prior studies have found negative effects of school transition from elementary schools to middle schools, which support our hypothesis on different relationships between reasoning and achievement at elementary and middle school levels. It is plausible to consider what differs in middle school environment from elementary schools (Theriot & Dupper, 2010). However, we argued that the differences are not only because students learn new science or mathematics contents across grade levels, but also because new academic demands and multiple sets of behavior rules are required as they grow up (Elias, 2001). Such new academic demands may include more complex reasoning skills or applying those skills to various contexts.

## Science Heuristic Writing Approach

Hand and Keys (1999) developed the SWH approach, which offers opportunities to pose questions, collect data, make claims based on evidence, and collectively construct and criticize their arguments. With the SWH approach, students use multiple modes of languages like pictures, graphs, equations, texts, and diagrams to present their ideas. Their ideas are structured with questions, claims, designs, and evidence, which leads to students' own arguments. Students pose their arguments in interactions with their peers and students have space to criticize and improve their arguments. At the completion of a curriculum unit, students are required to explain the big ideas of the units to younger students. Through multiple opportunities to express their understanding and engage in oral and written negotiations, students develop conceptual understanding of the concepts. The SWH approach promotes to utilize language – in its all forms –, argument, and dialogic interactions as epistemic tools simultaneously. The approach assures a natural environment for the learners, in which these epistemic tools are utilized strongly related to each other, develop together, and support each other. Because of the nature of science, argumentation is main characteristic of science communities, by extension, should be part of science learning environments. Cavagnetto (2010), in his comprehensive review on argument-based inquiry approaches in science education, asserts that immersive approaches are more promising for science learning and scientific literacy.

The SWH approach, as an immersive approach, has three phases: (1) The development of the underpinning epistemic framework phase, (2) the argument phase, and (3) the summary writing phase (Hand et al., 2017). The three phases do not occur in step by step; instead, there is a dependent-relational development between them. That is, with an initial introduction about the approach, students develop their epistemic framework. Then, students start to practice with argument phase and summary writing phase. They develop better framework as they use it. Within this cyclic process, the better framework students develop, the better practice students have in argument and summary writing; and vice versa.

The first phase, called development of the underpinning epistemic framework, pays attention "on the promotion of the 'big ideas' of the topic, the role of language, the role of negotiation as well as issues related to prior knowledge of the discipline, the structure and practice of argument, and the role of the group in learning" (Hand et al., 2017, p. 4). The second phase is the argument phase that requires to achieve to persuade oneself and others within a negotiation process. This phase includes more student talk and student argumentative writing utilizing question-claim-evidence structure to make sense about big ideas while working on their inquiry related to target concept. In the second phase, student have a transfer and translate from verbal language to written language –both include text and non-text representations, in other words utilization of multimodal representations. These translation and transitions require higher level cognition that maximize learning. Students are encouraged to complete an argumentative writing related to each concept under a unit. After finishing a unit, students are encouraged to write a summary about all the big ideas together. This is the third phase –the summary writing phase in which students go one step further of the argumentative writing. This phase requires "students to engage with language in determining what to represent (the content of the big idea) and how to represent (the types of modes used) the big ideas of the topic" (Hand et al., 2018, p. 5). This phase requires "complex cognitive processing" because the students need to engage with complex tasks such as selecting what to include, organizing ideas, translating science language to an audience appropriate language, choosing particular multimodal representations to explain the big idea accordingly, and transforming the constructed text into the target writing format. Moreover, one of the main differences of the summary writing with the SWH approach than the traditional summary writing is that students need to write to a person who does not know the topic. By doing so, students should be more illustrative in their writing; Instead of using big words, they use more explanatory words and non-text modes to make concepts clearer and more understandable.

Prior studies underscore the positive influences of the SWH approaches on development of reasoning as well as achievement in science and mathematics (Hand et al., 2018). Students who learn science with the SWH approach show more rapid development of reasoning abilities when compared to their peers in traditional classrooms (French et al., 2012). The implementation of the SWH approach contributes to better performance on standardized tests. In addition to gains in science achievement, when students learned sciences with the SWH approach at grade 5, they show higher achievement in mathematics at grades 5 and 7 (Hand et al., 2022).

Hand et al. (2022) also shows that students who learn science with the SWH approach in grade 5 demonstrated significant gains in inductive reasoning in mathematics.

Abductive reasoning in science was not observable in that study because of test specifications. Hand et al. found immediate positive influences on students' inductive/abductive reasoning. The SWH approach shows delayed effects on students' deductive reasoning in both mathematics and sciences at grade 7 when the students passed through the school transition from elementary to middle schools. This suggests that students could have gains in content knowledge and reasoning from learning science with the SWH approach. This study builds on Hand et al. (2022) and will examine students' interdependency of reasoning and content knowledge and the longitudinal development of the interdependency.

## III. RESEARCH QUESTIONS

The present research investigated the relationships between two critical resources students use within any learning environment – content knowledge and reasoning. For this purpose, we compared statistical models to answer the following main questions: (1) Is the relationships between students' reasoning and academic achievement non-linear when they learn science with the SWH approach? (2) Does students' background information influence on those relationships? (3) Are there differences in those relationships between learning environments – the SWH and traditional approaches? (4) Are those relationships stable across grades 5, 6, and 7? And (5) Is reasoning related to mathematics, science, and reading achievement differently?

## IV. METHODS

To address the research questions above, a three-stage research design was implemented: (1) the development of a model based upon a student-centered approach to learning called the Science Heuristic Writing (SWH) approach, (2) a replication phase to test the validity of the model across different learning environments, and (3) a stability phase to test the model with different datasets representing different learning environments. The supportive vector machine (SVM) was applied and compared to linear models.

### Supportive Vector Machine

We applied a supervised machine learning approach used for regression problems, for example, prediction of students' academic achievement scores. The machine learning approaches have been widely utilized to understanding data and finding patterns in many professional fields. Although this is deemed new, both machine learning and traditional statistics approaches are concerned with the same question, how we learn from data. However, there are several critical differences between machine learning and statistics that researchers should be aware of. Machine learning is mainly about predictions while statistics is about population, assumptions, and inferences. Because of this remarkable difference between them, machine learning requires no assumptions about variables like normality or linearity. Instead, one algorithm is applied to at least two datasets (training and scoring) to validate the output from the training dataset.

Among supervised machine learning approaches, we utilized a supportive vector machine (SVM) originally known as a strong classification technique among. An SVM deal with classification problems by finding an optimal hyperplane separating the sample to identify what we targeted. This technique is also applicable for regression problems. The radial basis function kernel was employed because the SVM with this kernel offer flexibility in modeling non-linear educational data. The radial kernel function is defined as

$$K(x, x') = \exp\left(-\sigma\left||x - x'|\right|^2\right)$$

where $\|x - x'\|^2$ represent the squared $L_2$ distance between two vectors $x$ and $x'$, and $\sigma$ is a free parameter. The larger σ is, the smaller value of the radial kernel function would be. This means that there are small differences in the values of the radial kernel function across the sample (in other words, the maximum and the minimum of this function in the sample are "close"). Then, if σ is large, it is likely to have a strict SVM decision boundary, which tends to overfit.

This research did not aim to find the best model comparing different kernels or other techniques like neural networks. Rather, we examined the performance of the SVM with the radial kernel to explain the relationships between critical thinking and academic achievement. Although the main purpose of machine learning is prediction, we argue that a model with repeated and successful prediction with multiple datasets can be a way to understand the interconnectedness of reasoning and content resources. This indicated that it is important to build a good model, which shows good-model-fit across multiple datasets (marked with a star Figure 1. Thus, we are cautious to identify overfitted models in this research. To avoid overfitting requires limiting the complexity of the models (regulation parameter C) and implementing five-fold cross-validation. Also, the best model is revealed via the hyperparameter optimization process.

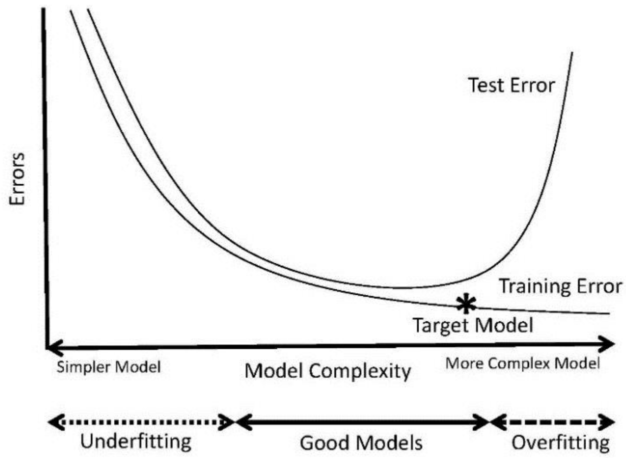


**Figure 1.** Complexity of models and model fits.

The R package caret (Kuhn, 2019) was utilized for model training and Parameter Tuning with a five-fold cross validation. As a result of hyperparameter optimization and cross-validation, the SVM models are evaluated with Root Mean Square Error (RMSE $= \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$ ), Mean Absolute Error ( MAE $= \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$ ), and $R^2 = \frac{\sum_{j=1}^{n}(\hat{y}_j - \bar{y})^2}{\sum_{j=1}^{n}(y_j - \bar{y})^2}$ where $y_j$ represent an observed score and $\hat{y}_j$ is a predicted score. Particularly, RMSE is used to select the optimal model to explain the relationships between cognitive resources and achievement, which is related to the average of Euclidean distances (L2 norm) between actual and predicted achievement scores. Using RMSE penalizes the large errors more than using MAE.

## Research Design

As discussed before, this research consists of the three sequential studies for the three different objectives. The three studies focused on prediction, replication, and stability as seen in Figure 2.

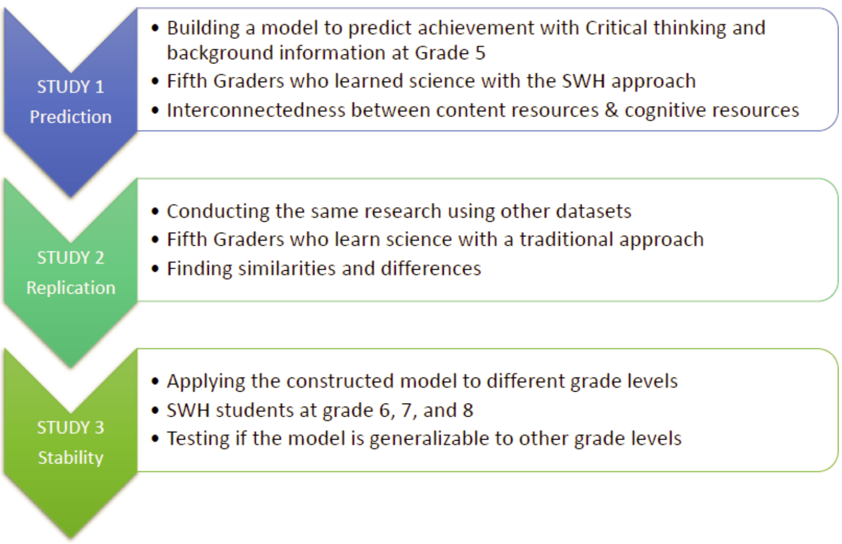


**Figure 2**. Three stages of the research

**Study 1 – predictive study.** The purpose of the study 1 is to build a statistical model to predict students' achievement in mathematics, science, and language using critical thinking scores, and other background information (gender, socioeconomic status index, etc.). The datasets included variables about fifth graders who learned science with the Science Heuristic Writing (SWH) approach. The model built in Study 1 is labeled as the SWH-G5 model.

To ensure validity of a dataset and the reliability of a model for further studies, the following procures was used: (1) preparing and cleaning data, (2) defining features of variables and re-scaling variables in the dataset, (3) applying SVM, (4) evaluating models based on RMSE, MAE, and $R^2$, and (5) interpreting results of students' predicted scores estimated by the best model. Specifically, we compared the model with only critical thinking scores as predictors and one with critical thinking scores and background information together. The results would indicate whether the relationships between critical thinking and academic achievement depend on student's background.

**Study 2 – replication study.** Although there have been many quantitative studies to build statistical models, only few replication studies using established models have been found (Cai et al., 2018). Replication studies are necessary to check generalizability of the models, which check whether the constructed model is over-fitted into the specific dataset in the previous studies. Replication studies allow us to apply one model to understand different student populations.

Study 2 is a replication of Study 1 using a different student population. The purpose of Study 2 is to examine the predictive power of the SWH-G5 model constructed in Study 1 with the data of fifth-grade student who have learned science in traditional classrooms. For this purpose, I will implement the following procedures: (1) preparing and cleaning data, (2) defining features of variables and re-scaling variables, (3) retrieving the SWH-G5 model from Study 1, (4) test the SWH-G5 model with the new dataset and comparing the model fit metrics with the results of Study 1, and (5) interpreting the results of model fit and students' predicted scores in the two different groups. If the model in the Study 1 is poorly performed with the data of students in the reference group, the procedure in Study 1 will be repeated to make a baseline model for the reference group, called the REF-G5 model. In addition, the results of the REF-G5 model were used to evaluate the performance of the SWH-G5 model with the dataset about students in traditional classrooms. This evaluation can inform how well the SWH-G5 model perform compared to the optimal model.

**Study 3 – stability study.** In Study 3, data from students different grade levels were analyzed utilizing the SWH-G5 model or the REF-G5 model (if it is significantly different from the SWH-G5 model) constructed through Studies 1 and 2. While Study 2 focused on generalizability of the SWH-G5 model between learning environments (SWH and traditional), the purpose of Study 3 is to test if the models at grade 5 is generalizable or stable across multiple grade levels, specifically to middle-school grade levels (grades 6, 7, and 8). This study 3 can also be considered as a replication study using the results of Studies 1 and 2, but I refer this study to a stability study to highlight the purpose of this study. Utilizing SVM, the research procedure in Study 3 is similar to the procedure used in Studies 1 and 2: (1) preparing and cleaning data, (2) defining features of variables and re-scaling variables, (3) retrieving the SWH-G5 model or the REF-G5 model from Studies 1 and 2, (4) evaluating the model with the data of different grade levels and compare the model fit metrics, and (5) interpret the results of model fit and students' predicted scores across grade levels. As done in Study 2, the optimal models at grades 6, 7, and 8 were constructed and utilized to evaluate how well the SWH-G5 model perform to predict achievement at grades 6, 7, and 8 in comparison with these optimal models – labeled as the G6 model, the G7 model, and the G8 model.

## Data Description

We utilized two existing datasets: one was collected in the random-control trial research funded by the Institute of Education Sciences (IES, the award number: R305A090094). This dataset includes fifth graders' performance in the Cornel critical thinking test and the Iowa Assessments, background information, and learning science approach (SWH or Traditional). The other dataset was collected in a research project funded by the Mathematics and Science Partnerships (MSP) program. This dataset also includes different students' performance in the Cornel critical thinking test and the Iowa Assessments at grades 6, 7, and 8.

**Table 1.** The number of students

| | | Grade 5 (SWH) | Grade 5 (Reference) | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Mathematics | Train | 1498 | 1232 | 877 | 944 | 778 |
| | Test | 735 | 604 | 429 | 462 | 380 |
| | Total | 2233 | 1836 | 1306 | 1406 | 1158 |
| Science | Train | 1498 | 1239 | 876 | 943 | 778 |
| | Test | 735 | 608 | 430 | 463 | 380 |
| | Total | 2233 | 1847 | 1306 | 1406 | 1158 |
| Reading | Train | 1499 | 1204 | 877 | 942 | 780 |
| | Test | 735 | 591 | 428 | 463 | 380 |
| | Total | 2234 | 1795 | 1305 | 1405 | 1160 |

The numbers of students are reported in Table 1. For the analysis with a machine learning approach, each dataset is split into two: we randomly assigned 67% of students into a train dataset and the others (33%) are distributed to a test dataset.

**Table 2.** Cornell critical thinking test

| Aspect of Critical Thinking | Items | # of Items | Definition |
|---|---|---|---|
| Induction | 3-25, 48, 50 | 25 | Ability to evaluate the facts that strengthen or weaken a given hypothesis |
| Deduction | 52-60, 67-76 | 24 | Ability to find a correct consequence from given antecedents |
| Observation | 27-50 | 24 | |
| Credibility | 27-50 | 24 | Ability to make judgments about whether, and to what extent, to believe someone else's assertion, usually in a situation in which the judger has no direct access to the basis for the assertion |
| Assumption Identification | 67-76 | 10 | Ability to fill a gap in a given reasoning statement |

### Selected Variables

We collected all the four sub-scores provided by the Cornel critical thinking test: Induction, deduction, observation/credibility, and assumption identification. The definition of each score is shown in Table 2. It should be noted that observation and credibility are considered as different constructs of critical thinking, but they are evaluated through the same items because it is difficult to examine those separately. In addition, the construct of assumption identification is examined using ten of the 24 items for deduction. This means that the construct of assumption identification is a subset of the construct of deduction.

All critical thinking scores are represented by the number of correct answers and we utilized national standard scores (mathematics, science, and reading) measured through the Iowa Assessments. Before data analysis, all variables about critical thinking and achievement were standardized. Table 3 shows all variables used in the data analysis including students' background information.

**Table 3.** Variables used in data analysis

| Model 1 | Model 2 | Description | |
|---|---|---|---|
| NSS_MT<br>NSS_SC<br>NSS_RT | NSS_MT<br>NSS_SC<br>NSS_RT | National Standard Scores provided by the Iowa Assessments | MT = Mathematics<br>SC = Science<br>RT = Reading |
| IND | IND | Induction scores in the CCTT | |
| DED | DED | Deduction scores in the CCTT | |
| OBS | OBS | Observation/Credibility scores in the CCTT | |
| ASM | ASM | Assumption Identification scores in the CCTT | |
| | SEX | Gender | 0 = Male, 1 = Female |
| | FRL | Free/reduced lunch | |
| | ELL | English language learners | |
| | AI | American Indian or Alaskan Native | |
| | ASN | Asian | |
| | BLK | Black or African American | |
| | HSP | Hispanic or Latino | |
| | HAW | Hawaiian/Pacific Islander | |
| | WHT | White | |
| | OTH | Other | |

## V. RESULTS

### Study 1

The results enable us to compare RMSE, $R^2$ and MAE of the models in the four ways: (1) between training and test datasets, (2) between SVM and linear models, (3) between Models 1 and 2, and (4) across mathematics, science, and reading. First, the two types of SWH models (linear and SVM) at grade 5 were built with the training set, and the models were evaluated with the test dataset. As seen in Tables 4, 5, and 6, the model fit results from each test dataset were slightly lower than the results from the corresponding training dataset. However, the differences of RMSE and MAE between the two different datasets were very small, which indicates that the established models were good models described in Figure 1, that is, neither underfitting nor overfitting.

Second, comparison between the SVM and linear models with each database showed similar model fits of these models. This result was found across the disciplines and the model types while showing that the values of RMSE ranges between 0.79 and 0.82 (see Tables 4, 5, and 6). Thus, the non-linear models – the SVM models – did not outperform the corresponding linear models. Third, it is interesting to see no goodness-of-fit improvement of Model 2 by adding variables about students' background information (ethnicity, ELL, and FRL) to Model 1. Thus, we did not have evidence for different relationships between critical thinking domains and achievement by students' backgrounds. Fourth, the similar model fits between the SVM and linear models as well as between Models 1 and 2 were found across mathematics, science, and reading. Therefore, the relationships between critical thinking domains and academic achievement did not depend on content areas.

**Table 4.** Model fit evaluation results (Mathematics)

| | | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|---|
| Type | Dataset | RMSE | $R^2$ | MAE | RMSE | $R^2$ | MAE |
| Linear | Train | 0.7995 | 0.3637 | 0.6420 | 0.7931 | 0.3738 | 0.6374 |
| | Test | 0.8142 | 0.3346 | 0.6500 | 0.8083 | 0.3441 | 0.6477 |
| SVM | Train | 0.7897 | 0.3796 | 0.6336 | 0.7923 | 0.3762 | 0.6353 |
| Radial | Test | 0.7955 | 0.3648 | 0.6340 | 0.7940 | 0.3672 | 0.6336 |

*Note*. σ = 0.02 and c = 0.5 for SVM model 1. σ = 0.01 and c = 0.25 for SVM model 2.

**Table 5**. Model fit evaluation results (Science)

| | | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|---|
| Type | Dataset | RMSE | $R^2$ | MAE | RMSE | $R^2$ | MAE |
| Linear | Train | 0.8084 | 0.3529 | 0.6439 | 0.8045 | 0.3598 | 0.6378 |
| | Test | 0.8159 | 0.3344 | 0.6528 | 0.8037 | 0.3542 | 0.6399 |
| SVM | Train | 0.7939 | 0.3729 | 0.6301 | 0.7948 | 0.3715 | 0.6301 |
| Radial | Test | 0.7924 | 0.3722 | 0.6299 | 0.7962 | 0.3663 | 0.6301 |

*Note*. σ = 0.02 and c = 0.5 for SVM model 1. σ = 0.01 and c = 0.25 for SVM model 2

**Table 6.** Model fit evaluation results (Reading)

| | | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|---|
| Type | Dataset | RMSE | $R^2$ | MAE | RMSE | $R^2$ | MAE |
| Linear | Train | 0.8154 | 0.3419 | 0.6343 | 0.8080 | 0.3539 | 0.6280 |
| | Test | 0.8010 | 0.3462 | 0.6191 | 0.7937 | 0.3580 | 0.6142 |
| SVM | Train | 0.7987 | 0.3696 | 0.6208 | 0.7978 | 0.3726 | 0.6208 |
| Radial | Test | 0.7816 | 0.3775 | 0.6048 | 0.7777 | 0.3837 | 0.6036 |

*Note*. σ = 0.05 and c = 0.5 for SVM model 1. σ = 0.01 and c = 0.75 for SVM model 2.

## Study 2

We construct the models of the SWH students in Study 1. The results showed that the SVM models showed similar performance with the linear models to predict academic achievement with critical thinking domains across mathematics, science, and reading. Furthermore, adding students' background variables did not improve the performance of the SVM model. In Study 2, we evaluated those linear and SVM models of the SWH students with reference data collected from the non-SWH students. The bold rows in Tables 7, 8, and 9 indicate how well each SWH model with the reference database perform to predict students' academic achievement with critical thinking domains. There were no significant differences between the results of the SWH students and the non-SWH students. The values of RMSE and MAE were even smaller with the reference database. These results were found in all mathematics, science, and readings. Therefore, the relationships between critical thinking domains and achievement showed by the SVM and linear models could be independent from learning environments – SWH and traditional.

**Table 7.** Evaluation of the SWH models with the non-SWH database (Mathematics)

| Type (kernel) | Dataset | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|---|
| | | RMSE | $R^2$ | MAE | RMSE | $R^2$ | MAE |
| Linear | SWH Train | 0.7995 | 0.3637 | 0.6420 | 0.7931 | 0.3738 | 0.6374 |
| | SWH Test | 0.8142 | 0.3346 | 0.6500 | 0.8083 | 0.3441 | 0.6477 |
| | Non-SWH Test | 0.7844 | 0.3847 | 0.6308 | 0.7820 | 0.3885 | 0.6256 |
| SVM Radial | SWH Train | 0.7897 | 0.3796 | 0.6336 | 0.7923 | 0.3762 | 0.6353 |
| | SWH Test | 0.7955 | 0.3648 | 0.6340 | 0.7940 | 0.3672 | 0.6336 |
| | Non-SWH Test | 0.7775 | 0.3954 | 0.6266 | 0.7803 | 0.3911 | 0.6275 |

**Table 8.** Evaluation of the SWH models with the non-SWH database (Science)

| Type (kernel) | Dataset | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|---|
| | | RMSE | $R^2$ | MAE | RMSE | $R^2$ | MAE |
| Linear | SWH Train | 0.8084 | 0.3529 | 0.6439 | 0.8045 | 0.3598 | 0.6378 |
| | SWH Test | 0.8159 | 0.3344 | 0.6528 | 0.8037 | 0.3542 | 0.6399 |
| | Non-SWH Test | 0.7805 | 0.3909 | 0.6306 | 0.7758 | 0.3982 | 0.6276 |
| SVM Radial | SWH Train | 0.7939 | 0.3729 | 0.6301 | 0.7948 | 0.3715 | 0.6301 |
| | SWH Test | 0.7924 | 0.3722 | 0.6299 | 0.7962 | 0.3663 | 0.6301 |
| | Non-SWH Test | 0.7712 | 0.4053 | 0.6289 | 0.7688 | 0.4089 | 0.6147 |

**Table 9.** Evaluation of the SWH models with the non-SWH database (Reading)

| Type (kernel) | Dataset | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|---|
| | | RMSE | $R^2$ | MAE | RMSE | $R^2$ | MAE |
| Linear | SWH Train | 0.8154 | 0.3419 | 0.6343 | 0.8080 | 0.3539 | 0.6280 |
| | SWH Test | 0.8010 | 0.3462 | 0.6191 | 0.7937 | 0.3580 | 0.6142 |
| | Non-SWH Test | 0.7681 | 0.4100 | 0.6017 | 0.7649 | 0.4149 | 0.5977 |
| SVM Radial | SWH Train | 0.7987 | 0.3696 | 0.6208 | 0.7978 | 0.3726 | 0.6208 |
| | SWH Test | 0.7816 | 0.3775 | 0.6048 | 0.7777 | 0.3837 | 0.6036 |
| | Non-SWH Test | 0.7520 | 0.4345 | 0.5897 | 0.7534 | 0.5882 | 0.4323 |

We found that the SWH models showed similar model fits when they applied to the non-SWH database. After evaluating these SWH models, we wondered how different the SWH models are from the best models built with the non-SWH database. In Tables 10, 11, and 12, the bold rows were the model fit results of the SWH models with the non-SWH datasets from Tables 7, 8, and 9. The results in Tables 10, 11, and 12 showed that the SWH model is indifferent from the best model of each dataset. Even Models 1 and 2 showed the same patterns. The results are consistent across mathematics, science, and readings. Therefore, we concluded that each of the linear or non-linear relationships between critical thinking domains and achievement independent from students' backgrounds, learning environments, and content areas.

**Table 10.** Model fits of best model built with the non-SWH dataset in comparison with the SWH model (Mathematics)

| | | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|---|
| Type | Dataset | RMSE | $R^2$ | MAE | RMSE | $R^2$ | MAE |
| | Non-SWH Test | 0.7844 | 0.3847 | 0.6308 | 0.7820 | 0.3885 | 0.6256 |
| Linear | Train | 0.7874 | 0.3780 | 0.6359 | 0.7818 | 0.3871 | 0.6305 |
| | Test | 0.7808 | 0.3950 | 0.6241 | 0.7818 | 0.3934 | 0.6232 |
| | Non-SWH Test | 0.7775 | 0.3954 | 0.6266 | 0.7803 | 0.3911 | 0.6275 |
| SVM | Train | 0.7776 | 0.3946 | 0.6291 | 0.7771 | 0.3955 | 0.6305 |
| Radial | Test | 0.7721 | 0.4084 | 0.6184 | 0.7751 | 0.4037 | 0.6191 |

*Note.* σ = 0.05 and c = 0.25 for SVM model 1. σ = 0.01 and c = 0.25 for SVM model 2.

**Table 11.** Model fits of best model built with the non-SWH dataset in comparison with the SWH model (Science)

| | | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|---|
| Type | Dataset | RMSE | $R^2$ | MAE | RMSE | $R^2$ | MAE |
| | Non-SWH Test | 0.7805 | 0.3909 | 0.6306 | 0.7758 | 0.3982 | 0.6276 |
| Linear | Train | 0.7870 | 0.3770 | 0.6352 | 0.7858 | 0.3790 | 0.6328 |
| | Test | 0.7717 | 0.4144 | 0.6273 | 0.7650 | 0.4245 | 0.6246 |
| | Non-SWH Test | 0.7712 | 0.4053 | 0.6289 | 0.7688 | 0.4089 | 0.6147 |
| SVM | Train | 0.7722 | 0.3996 | 0.6199 | 0.7719 | 0.4013 | 0.6160 |
| Radial | Test | 0.7599 | 0.4322 | 0.6122 | 0.7592 | 0.4333 | 0.6123 |

*Note.* σ = 0.04 and c = 0.25 for SVM model 1. σ = 0.01 and c = 0.25 for SVM model 2.

**Table 12**. Model fits of best model built with the non-SWH dataset in comparison with the SWH model (Reading)

| | | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|---|
| Type | Dataset | RMSE | $R^2$ | MAE | RMSE | $R^2$ | MAE |
| | Non-SWH Test | 0.7681 | 0.4100 | 0.6017 | 0.7649 | 0.4149 | 0.5977 |
| Linear | Train | 0.7630 | 0.4219 | 0.6018 | 0.7605 | 0.4248 | 0.5996 |
| | Test | 0.7769 | 0.3949 | 0.5998 | 0.7689 | 0.4073 | 0.5897 |
| | Non-SWH Test | 0.7520 | 0.4345 | 0.5897 | 0.7534 | 0.5882 | 0.4323 |
| SVM | Train | 0.7435 | 0.4489 | 0.5876 | 0.7425 | 0.4514 | 0.5860 |
| Radial | Test | 0.7591 | 0.4223 | 0.5864 | 0.7501 | 0.4360 | 0.5785 |

*Note.* σ = 0.075 and c = 0.25 for SVM model 1. σ = 0.01 and c = 1.5 for SVM model 2.

## Study 3

In Study 1, we constructed the SVM models (Both Models 1 and 2) of fifth graders who learned science with the SWH approach in order to predict achievement using critical thinking domains. We found that the SVM models at Grade 5 showed similar predictive power with the corresponding linear models. In Study 2, the linear and SVM models of SWH students can also predict achievement of non-SWH students at Grade 5. Then, in Study 3, we tested these SWH models with datasets of students at other grade levels.

Tables 13, 14, and 15 show that the SWH models (both Models 1 and 2) did not have the same degree of predictive power at grades 6, 7, and 8. The values of RMSE increased to around 0.85. The $R^2$ index showed more notable changes. On the one hand, the values of $R^2$ were above 0.3 when the SWH models were tested with the fifth-grade datasets. On the other hand, the values of $R^2$ were less than 0.3 (mostly less than 0.25). These findings were identical across mathematics, science, and reading. However, it is necessary to compare the best models at each grade level and the corresponding SWH models. This is because, it could be the best results that we can get from the databases of grades 6, 7, and 8 although the model fits were worsen.

Tables 16, 17, and 18 show the model fit results of the best models (both Models 1 and 2) at each grade level. The results showed that the best models can predict students' achievement better than the SVM models. This means that stability of the SVM model at grade 5 was challenged. Although the SVM model can be replicated across learning environments, different grade levels could indicate different relationships between critical thinking domains and academic achievement. It should be noted that there was so remarkable difference in the results among mathematics, science, and readings.

**Table 13.** Evaluation of the SWH models with the databases of other grade levels (Mathematics)

| Type (kernel) | Dataset | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|---|
| | | RMSE | $R^2$ | MAE | RMSE | $R^2$ | MAE |
| Linear | SWH Train | 0.7995 | 0.3637 | 0.6420 | 0.7931 | 0.3738 | 0.6374 |
| | SWH Test | 0.8142 | 0.3346 | 0.6500 | 0.8083 | 0.3441 | 0.6477 |
| | Non-SWH Test | 0.7844 | 0.3847 | 0.6308 | 0.7820 | 0.3885 | 0.6256 |
| | Grade 6 Test | 0.8880 | 0.2042 | 0.7093 | 0.8744 | 0.2284 | 0.6946 |
| | Grade 7 Test | 0.8652 | 0.2276 | 0.6955 | 0.8528 | 0.2497 | 0.6821 |
| | Grade 8 Test | 0.8949 | 0.1764 | 0.7127 | 0.8846 | 0.1952 | 0.7034 |
| SVM Radial | SWH Train | 0.7897 | 0.3796 | 0.6336 | 0.7923 | 0.3762 | 0.6353 |
| | SWH Test | 0.7955 | 0.3648 | 0.6340 | 0.7940 | 0.3672 | 0.6336 |
| | Non-SWH Test | 0.7775 | 0.3954 | 0.6266 | 0.7803 | 0.3911 | 0.6275 |
| | Grade 6 Test | 0.8684 | 0.2389 | 0.6903 | 0.8699 | 0.2363 | 0.6878 |
| | Grade 7 Test | 0.8392 | 0.2733 | 0.6759 | 0.8395 | 0.2729 | 0.6756 |
| | Grade 8 Test | 0.8835 | 0.1973 | 0.7041 | 0.8920 | 0.1817 | 0.7093 |

*Note.* $\sigma = 0.02$ and $c = 0.5$ for SVM model 1. $\sigma = 0.01$ and $c = 0.25$ for SVM model 2.

**Table 14.** Evaluation of the SWH models with the databases of other grade levels (Science)

| Type (kernel) | Dataset | Model 1 RMSE | Model 1 $R^2$ | Model 1 MAE | Model 2 RMSE | Model 2 $R^2$ | Model 2 MAE |
|---|---|---|---|---|---|---|---|
| Linear | SWH Train | 0.8084 | 0.3529 | 0.6439 | 0.8045 | 0.3598 | 0.6378 |
| | SWH Test | 0.8159 | 0.3344 | 0.6528 | 0.8037 | 0.3542 | 0.6399 |
| | Non-SWH Test | 0.7805 | 0.3909 | 0.6306 | 0.7758 | 0.3982 | 0.6276 |
| | Grade 6 Test | 0.8705 | 0.2323 | 0.6881 | 0.8598 | 0.2510 | 0.6782 |
| | Grade 7 Test | 0.8725 | 0.2031 | 0.6796 | 0.8680 | 0.2113 | 0.6762 |
| | Grade 8 Test | 0.8749 | 0.2369 | 0.7001 | 0.8594 | 0.2638 | 0.6848 |
| SVM Radial | SWH Train | 0.7939 | 0.3729 | 0.6301 | 0.7948 | 0.3715 | 0.6301 |
| | SWH Test | 0.7924 | 0.3722 | 0.6299 | 0.7962 | 0.3663 | 0.6301 |
| | Non-SWH Test | 0.7712 | 0.4053 | 0.6289 | 0.7688 | 0.4089 | 0.6147 |
| | Grade 6 Test | 0.8680 | 0.2367 | 0.6843 | 0.8750 | 0.2242 | 0.6902 |
| | Grade 7 Test | 0.8693 | 0.2089 | 0.6779 | 0.8830 | 0.1838 | 0.6898 |
| | Grade 8 Test | 0.8750 | 0.2367 | 0.6976 | 0.8987 | 0.1949 | 0.7192 |

*Note*. σ = 0.02 and c = 0.5 for SVM model 1. σ = 0.01 and c = 0.25 for SVM model 2.

**Table 15.** Evaluation of the SWH models with the databases of other grade levels (Reading)

| Type (kernel) | Dataset | Model 1 RMSE | Model 1 $R^2$ | Model 1 MAE | Model 2 RMSE | Model 2 $R^2$ | Model 2 MAE |
|---|---|---|---|---|---|---|---|
| Linear | SWH Train | 0.8154 | 0.3419 | 0.6343 | 0.8080 | 0.3539 | 0.6280 |
| | SWH Test | 0.8010 | 0.3462 | 0.6191 | 0.7937 | 0.3580 | 0.6142 |
| | Non-SWH Test | 0.7681 | 0.4100 | 0.6017 | 0.7649 | 0.4149 | 0.5977 |
| | Grade 6 Test | 0.8734 | 0.2250 | 0.6973 | 0.8723 | 0.2269 | 0.6933 |
| | Grade 7 Test | 0.8464 | 0.2474 | 0.6819 | 0.8474 | 0.2455 | 0.6845 |
| | Grade 8 Test | 0.8629 | 0.2406 | 0.6928 | 0.8551 | 0.2542 | 0.6848 |
| SVM Radial | SWH Train | 0.7987 | 0.3696 | 0.6208 | 0.7978 | 0.3726 | 0.6208 |
| | SWH Test | 0.7816 | 0.3775 | 0.6048 | 0.7777 | 0.3837 | 0.6036 |
| | Non-SWH Test | 0.7520 | 0.4345 | 0.5897 | 0.7534 | 0.5882 | 0.4323 |
| | Grade 6 Test | 0.8641 | 0.2414 | 0.6894 | 0.8787 | 0.2156 | 0.6964 |
| | Grade 7 Test | 0.8271 | 0.2811 | 0.6656 | 0.8450 | 0.2497 | 0.6783 |
| | Grade 8 Test | 0.8524 | 0.2590 | 0.6793 | 0.8706 | 0.2269 | 0.6933 |

*Note*. σ = 0.05 and c = 0.5 for SVM model 1. σ = 0.01 and c = 0.75 for SVM model 2.

**Table 16.** Model fits of best models built with the datasets of grades 6, 7, and 8 in comparison to the test results of the SWH model (Mathematics)

| | | | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|---|---|
| | Type | Dataset | RMSE | $R^2$ | MAE | RMSE | $R^2$ | MAE |
| Grade 6[a] | Linear | Grade 6 Test | 0.8880 | 0.2042 | 0.7093 | 0.8744 | 0.2284 | 0.6946 |
| | | Train | 0.8523 | 0.2614 | 0.6862 | 0.8363 | 0.2894 | 0.6683 |
| | | Test | 0.8183 | 0.3375 | 0.6510 | 0.8035 | 0.3613 | 0.6411 |
| | SVM Radial | Grade 6 Test | 0.8684 | 0.2389 | 0.6903 | 0.8699 | 0.2363 | 0.6878 |
| | | Train | 0.8509 | 0.2701 | 0.6760 | 0.8447 | 0.2849 | 0.6686 |
| | | Test | 0.8192 | 0.3361 | 0.6467 | 0.8183 | 0.3376 | 0.6427 |
| Grade 7[b] | Linear | Grade 7 Test | 0.8652 | 0.2276 | 0.6955 | 0.8528 | 0.2497 | 0.6821 |
| | | Train | 0.8277 | 0.2921 | 0.6659 | 0.8123 | 0.3179 | 0.6486 |
| | | Test | 0.8068 | 0.3372 | 0.6574 | 0.7896 | 0.3652 | 0.6356 |
| | SVM Radial | Grade 7 Test | 0.8392 | 0.2733 | 0.6759 | 0.8395 | 0.2729 | 0.6756 |
| | | Train | 0.8210 | 0.3027 | 0.6584 | 0.8208 | 0.3044 | 0.6579 |
| | | Test | 0.8118 | 0.3289 | 0.6572 | 0.7986 | 0.3505 | 0.6414 |
| Grade 8[c] | Linear | Grade 8 Test | 0.8949 | 0.1764 | 0.7127 | 0.8846 | 0.1952 | 0.7034 |
| | | Train | 0.8364 | 0.2969 | 0.6669 | 0.8293 | 0.3115 | 0.6568 |
| | | Test | 0.8172 | 0.2827 | 0.6456 | 0.8099 | 0.2954 | 0.6303 |
| | SVM Radial | Grade 8 Test | 0.8835 | 0.1973 | 0.7041 | 0.8920 | 0.1817 | 0.7093 |
| | | Train | 0.8214 | 0.3248 | 0.6466 | 0.8167 | 0.3336 | 0.6430 |
| | | Test | 0.8382 | 0.2453 | 0.6460 | 0.8230 | 0.2725 | 0.6330 |

[a] $\sigma = 0.03$ and $c = 0.5$ for SVM model 1. $\sigma = 0.01$ and $c = 1$ for SVM model 2.
[b] $\sigma = 0.03$ and $c = 0.25$ for SVM model 1. $\sigma = 0.01$ and $c = 0.5$ for SVM model 2.
[c] $\sigma = 0.075$ and $c = 1$ for SVM model 1. $\sigma = 0.01$ and $c = 1.5$ for SVM model 2.

**Table 17.** Model fits of best models built with the datasets of grades 6, 7, and 8 in comparison to the test results of the SWH model (Science)

| | | | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|---|---|
| | Type | Dataset | RMSE | $R^2$ | MAE | RMSE | $R^2$ | MAE |
| Grade 6[a] | Linear | Grade 6 Test | 0.8705 | 0.2323 | 0.6881 | 0.8598 | 0.2510 | 0.6782 |
| | | Train | 0.8443 | 0.2698 | 0.6725 | 0.8341 | 0.2869 | 0.6607 |
| | | Test | 0.8395 | 0.3085 | 0.6684 | 0.8256 | 0.3312 | 0.6542 |
| | SVM Radial | Grade 6 Test | 0.8680 | 0.2367 | 0.6843 | 0.8750 | 0.2242 | 0.6902 |
| | | Train | 0.8387 | 0.2838 | 0.6642 | 0.8359 | 0.2874 | 0.6590 |
| | | Test | 0.8417 | 0.3048 | 0.6607 | 0.8383 | 0.3104 | 0.6590 |
| Grade 7[b] | Linear | Grade 7 Test | 0.8725 | 0.2031 | 0.6796 | 0.8680 | 0.2113 | 0.6762 |
| | | Train | 0.7955 | 0.3228 | 0.6319 | 0.7837 | 0.3425 | 0.6206 |
| | | Test | 0.8765 | 0.2438 | 0.6637 | 0.8757 | 0.2451 | 0.6722 |
| | SVM Radial | Grade 7 Test | 0.8693 | 0.2089 | 0.6779 | 0.8830 | 0.1838 | 0.6898 |
| | | Train | 0.7956 | 0.3237 | 0.6337 | 0.7942 | 0.3284 | 0.6269 |
| | | Test | 0.8733 | 0.2493 | 0.6633 | 0.8847 | 0.2296 | 0.6776 |
| Grade 8[c] | Linear | Grade 8 Test | 0.8749 | 0.2369 | 0.7001 | 0.8594 | 0.2638 | 0.6848 |
| | | Train | 0.8265 | 0.3475 | 0.6634 | 0.8187 | 0.3589 | 0.6549 |
| | | Test | 0.7895 | 0.3298 | 0.6308 | 0.7701 | 0.3624 | 0.6226 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | SVM Radial | Grade 8 Test | 0.8750 | 0.2367 | 0.6976 | 0.8987 | 0.1949 | 0.7192 |
| | | Train | 0.8157 | 0.3693 | 0.6467 | 0.8115 | 0.3751 | 0.6391 |
| | | Test | 0.7975 | 0.3161 | 0.6353 | 0.7936 | 0.3228 | 0.6331 |

[a] σ = 0.01 and c = 10 for SVM model 1. σ = 0.01 and c = 0.75 for SVM model 2.
[b] σ = 0.01 and c = 0.25 for SVM model 1. σ = 0.02 and c = 0.5 for SVM model 2.
[c] σ = 0.05 and c = 0.75 for SVM model 1. σ = 0.01 and c = 1.5 for SVM model 2.

**Table 18.** Model fits of best models built with the datasets of grades 6, 7, and 8 in comparison to the test results of the SWH model (Reading)

| | | | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|---|---|
| | Type | Dataset | RMSE | $R^2$ | MAE | RMSE | $R^2$ | MAE |
| Grade 6[a] | Linear | Grade 6 Test | 0.8734 | 0.2250 | 0.6973 | 0.8723 | 0.2269 | 0.6933 |
| | | Train | 0.8120 | 0.3334 | 0.6471 | 0.8094 | 0.3378 | 0.6458 |
| | | Test | 0.8221 | 0.3106 | 0.6564 | 0.8190 | 0.6470 | 0.3159 |
| | SVM Radial | Grade 6 Test | 0.8641 | 0.2414 | 0.6894 | 0.8787 | 0.2156 | 0.6964 |
| | | Train | 0.8068 | 0.3500 | 0.6374 | 0.8017 | 0.3576 | 0.6314 |
| | | Test | 0.8153 | 0.3220 | 0.6424 | 0.8060 | 0.3373 | 0.6289 |
| Grade 7[b] | Linear | Grade 7 Test | 0.8464 | 0.2474 | 0.6819 | 0.8474 | 0.2455 | 0.6845 |
| | | Train | 0.7897 | 0.3461 | 0.6404 | 0.7887 | 0.3475 | 0.6421 |
| | | Test | 0.7804 | 0.3648 | 0.6267 | 0.7773 | 0.3699 | 0.6237 |
| | SVM Radial | Grade 7 Test | 0.8271 | 0.2811 | 0.6656 | 0.8450 | 0.2497 | 0.6783 |
| | | Train | 0.7862 | 0.3551 | 0.6298 | 0.7888 | 0.3491 | 0.6301 |
| | | Test | 0.7826 | 0.3613 | 0.6232 | 0.7872 | 0.3537 | 0.6271 |
| Grade 8[c] | Linear | Grade 8 Test | 0.8629 | 0.2406 | 0.6928 | 0.8551 | 0.2542 | 0.6848 |
| | | Train | 0.7816 | 0.3734 | 0.6282 | 0.7888 | 0.3626 | 0.6345 |
| | | Test | 0.8000 | 0.3594 | 0.6486 | 0.7982 | 0.3622 | 0.6450 |
| | SVM Radial | Grade 8 Test | 0.8524 | 0.2590 | 0.6793 | 0.8706 | 0.2269 | 0.6933 |
| | | Train | 0.7778 | 0.3820 | 0.6251 | 0.7773 | 0.3823 | 0.6244 |
| | | Test | 0.8125 | 0.3392 | 0.6546 | 0.8124 | 0.3393 | 0.6541 |

[a] σ = 0.1 and c = 2 for SVM model 1. σ = 0.01 and c = 2 for SVM model 2.
[b] σ = 0.02 and c = 0.75 for SVM model 1. σ = 0.01 and c = 0.5 for SVM model 2.
[c] σ = 0.075 and c = 0.25 for SVM model 1. σ = 0.01 and c = 0.25 for SVM model 2.

## VI. DISCUSSION AND CONCLUSION

In this study, we generated models about interdependency between reasoning and content. First, we analyzed the data about the SWH students to build the linear and SVM models representing relationships between critical thinking domains and academic achievement at grade 5. Then, the SWH models were applied to the non-SWH data at the same grade level in order to examine if the models were replicated. Lastly, the SWH models were evaluated with the data at grades 6, 7, and 8 so that the relationships represented by the SWH models were stable across the grade levels.

The findings from the three stages of analysis provided multiple discussion themes by model comparison. First, comparison between the SVM and linear models can answer if the relationships between critical thinking domains and academic achievement are complex beyond linearity. Second, evaluating Models 1 and 2 can show if the relationships differ by students' background. Third, comparison between SWH and non-SWH models at grade 5 can indicate whether the relationships depend on learning environments. Fourth, we can investigate if the relationships differ by content area (mathematics, science, and reading). Lastly, stability of the relationships across grades 6, 7, and 8 can be assessed applying the models at grade 5.

There are five themes to discuss about the relationship between critical thinking domains and academic achievement: (1) linearity, (2) no influence of background information on the relationships, (3) no difference between learning environments, (4) different relationships among the grade levels, and (5) similar patterns across mathematics, science, and reading.

First, the findings showed that the SVM models did not outperform the corresponding linear models across all observed cases. This finding showed that the non-linear models were not able to identify evidence for complex relationships between critical thinking domains and academic achievements beyond linearity. Thus, this result seems aligned to the idea that the better thinker, the higher achiever. However, we do not conclude that the relationships are linear, but rather we argued that more studies are required to answer the following question: Is it possible that the complexity oversimplified due to the nature of measurement? Although the relationships were theoretically argued as complex, but such empirical relationships cannot be discussed separately from how the variables were defined and measured.

Second, when we compared Models 1 and 2 regardless of linearity of the relationships, Model 1 showed similar predictive power with Model 2. This means that adding variables about students' background information (socioeconomic status, English Language Learner, and ethnicity groups) cannot improve predicting academic achievement with critical thinking scores. This finding indicates that the relationships between critical thinking and achievement are identical among students' different backgrounds. Then, we questioned if students in minoritized groups (e.g., low SES families and ELL) can be successful when they develop critical thinking skills. While educators have suggested various ways to help students in underrepresented groups, this finding suggests that learning environments promoting critical thinking also help underrepresented students improve their achievement.

Third, there was no difference between SWH and non-SWH models at grade 5. This indicates that the relationships between critical thinking and academic achievement are independent from learning environments. This finding helps us better understand the finding of Hand et al. (2018): the SWH approach improved students' critical thinking, mathematics and reading achievement. A standardized assessment is a distant tool to evaluate what students develop through SWH instructions. Thus, the SWH approach contributed to students' higher critical thinking scores, which resulted in higher achievement of the SWH students although the relationships between critical thinking and achievement were consistent between the learning environments. However, further research is still required to collect more empirical evidence about how learning environments change the relationships between critical thinking and achievement. Because machine learning is a technique to build a generalizable model, a difference between the SWH and non-SWH models was possible ignored. We will discuss this measurement issue at the end of the discussion section.

For the fourth and fifth points, the relationships of critical thinking and achievement at grade 5, identified in the SWH models, were not stable after school transition to middle schools. Furthermore, the results at grade 8 showed different patterns across mathematics, science, and reading in comparison to the other grade levels. This indicates that there would be interaction between grade levels and disciplines, which changes the connectedness between critical thinking and achievement across different grade levels. However, it is uncertain what variables would be confounding in the findings.

This research applied a new approach to understand the relationships between critical thinking and achievement scores. A remarkable difference of machine learning approaches from traditional statistics is that machine learning approaches do not require assumptions about variables like normality or linearity. Instead, because the purpose of machine learning is prediction, one algorithm is applied to training and test datasets to validate the models built with the training dataset. Thus, we were not able to argue statistical significance of differences among the models in this study. However, we found that there is no difference in prediction power of linear and non-linear models. Furthermore, the findings showed that the models of the SWH and non-SWH students were similar regardless of linearity of the relationships between critical thinking and achievement.

Applying the SVM models in this study provided additional evidence for validity of the Iowa assessments. The Iowa Assessments (2015) provide correlations between scores on the Iowa Assessments and scores on Cognitive Abilities Tests for concurrent validity coefficients. In addition to this traditional type of statistics, accurate predictions of Iowa assessment scores using the CCTT scores could be additional evidence of concurrent validity of the Iowa Assessments. In other words, cognitive practices have potential to predict students' achievement, which means that academic achievement and critical thinking – a collective of cognitive practices – are related. Furthermore, this relationships between critical thinking and achievement are identical across different learning environments, grade levels, and students' background.

Another measurement issue is about the idea of power. With the traditional perspective, power of hypothesis tests means probability to reject the null hypothesis. In other words, it is usually about how sensitive applied statistics were to differences of statistical models. However, in machine learning, the predictive power of models means that the models are applicable to various contexts. Thus, a generalizable machine learning model should not be sensitive. A machine learning approach for prediction might be inappropriate to identify differences between models or groups. However, it should be noted that we recognized differences among the models at different grade levels. Thus, we suggest answering researcher's questions with multiple quantitative methods although research methods should be corresponding to research questions.

This research did not provide evidence for non-linear relationships between critical thinking and academic achievement although SVM models have been studied as a powerful tool for prediction. We applied one specific model to our data, which means that it should be very cautious to conclude linearity. It is required to test various non-linear regression models (for example, neural network) for better understanding of the relationships between critical thinking and academic achievement. Furthermore, it should be cautious in generalizing non-linear relationships between critical thinking and achievement. Someone can support linearity considering that the better thinker, the higher achiever. However, replication studies are required for better understanding of those relationships.

Many schools have noted that critical thinking and reasoning are what their students should develop. Simultaneously, standardized tests have significant influences on how students learn in their schools. This study helps schools understand the relationships between critical thinking and achievement better, as well as support their students promote both reasoning and performance on standardized tests. Because of strengths of a machine learning approach, the models built in this study can be easily distributed to schools and teachers and utilized to predict students' results in summative assessments with their reasoning skills developed in classrooms.

## References

Akkus, R., Gunel, M., & Hand, B. (2007). Comparing an inquiry- based approach known as the science writing heuristic to traditional science teaching practices: Are there differences?. *International Journal of Science Education, 29*(14), 1745–1765.

Azar, A. (2010). The effect of critical thinking dispositions on students achievement in selection and placement exam for university in Turkey. *Journal of Turkish Science Education, 7*(1), 61–73.

Bailin, S. (2002). Critical thinking and science education. *Science & Education, 11*, 361–375.

Blair, J. A. (2016). An early exchange on the interpretation of arguments in texts. *Informal Logic, 36*(1), 83–91.

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., & Hiebert, J. (2018). The role of replication studies in educational research. *Journal for Research in Mathematics Education, 49*(1), 2–8.

Cavagnetto, A. R. (2010). Argument to foster scientific literacy: A review of argument interventions in K–12 science contexts. *Review of Educational Research, 80*(3), 336–371. doi:10.3102/0034654310376953

Cleophas, T. J., & Zwinderman, A. H. (2013). Support vector machines. In *Machine learning in medicine* (pp. 155–161). Springer.

Elias, M. J. (2001). Middle school transition: It's harder than you think. Making the transition to middle school successful. *Middle Matters*, 1–2. https://resources.finalsite.net/images/v1545339111/onteorak12nyus/puho8gytny9d0eijwixl/MSTransitionHarder.pdf

Ennis, R. H., Millman, J., & Tomko, T. N. (2004). *Cornell critical thinking tests level X & level Z manual* (4th Ed.). The Critical Thinking Co.

French, B. F., Hand, B., Therrien, W. J., & Valdivia Vazquez, J. A. (2012). Detection of sex differential item functioning in the Cornell Critical Thinking Test. *European Journal of Psychological Assessment, 28*(3), 201–207. doi:10.1027/1015-5759/a000127

Hand, B., Choi, K., & Hwang, J. (2022). *A longitudinal study of reasoning development in mathematics and science through argument-based inquiry*. The manuscript in preparation.

Hand, B., & Keys, C. W. (1999). Inquiry investigation. *Science Teacher, 66*(4), 27–29.

Hand, B., Norton-Meier, L., & Jang, J. Y. (Eds.). (2017). *More voices from the classroom: International teachers' experience with argument-based inquiry*. Springer

Hand, B., Shelley, M. C., Laugerman, M., Fostvedt, L., & Therrien, W. (2018). Improving critical thinking growth for disadvantaged groups within elementary school science: A randomized controlled trial using the Science Writing Heuristic approach. *Science Education, 102*(4), 693–710. doi:10.1002/sce.21341

Hyman, L. C. (2015). *Middle school plunge: A mixed-methods study exploring 6th grade students' perceptions of their transition to middle school experiences and its influences on school achievement and performance for urban youth.* (Unpublished doctoral dissertation). Rowan University.

Khomenko, I. (2016). Types of reasoning in argumentation. In D. Mohammed & M. Lewiński (Eds.), *Argumentation and reasoned action: Proceedings of the 1st European conference on Argumentation, Lisbon, 2015* (Vol. 2, pp. 617–627). College Publication.

Kingir, S., Geban, O., & Gunel, M. (2012). How does the science writing heuristic approach affect students' performances of different academic achievement levels? A case for high school chemistry. *Chemistry Education Research and Practice, 13*(4), 428–436.

Kirwan, C. (1995). Reasoning. In T. Honderich (Ed.), *The Oxford companion to philosophy* (2nd ed., pp. 791). Oxford University Press.

Kuhn, M. (2019). Caret: Classification and regression training. R package version 6.0-84.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Lawson, A. E. (2005). What is the role of induction and deduction in reasoning and scientific inquiry? *Journal of Research in Science Teaching, 42*(6), 716–740.

Lunenburg, F. C. (2011). Critical thinking and constructivism techniques for improving student achievement. *National Forum of Teacher Education Journal, 21*(3), 1–9.

Moshman, D. (2015). *Epistemic cognition and development: The psychology of justification and truth*. Psychology Press.

National Council of Teachers of Mathematics [NCTM]. (2000). *Principles and standards for school mathematics*. NCTM.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). *Common core state standards for mathematics.* http://www.corestandards.org/Math/

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press.

Sadler, T. D., & Zeidler, D. L. (2005). The significance of content knowledge for informal reasoning regarding socioscientific issues: Applying genetics knowledge to genetic engineering issues. *Science Education, 89*(1), 71–93. doi:10.1002/sce.20023

Smith, O. (1953). The improvement of critical thinking. *Progressive Education, 30*(5), 129–134.

Stephenson, N. S., & Sadler-McKnight, N. P. (2016). Developing critical thinking skills using the science writing heuristic in the chemistry laboratory. *Chemistry Education Research and Practice, 17*(1), 72–79.

Stupnisky, R. H., Renaud, R. D., Daniels, L. M., Haynes, T. L., & Perry, R. P. (2008). The interrelation of first-year college students' critical thinking disposition, perceived academic control, and academic achievement. *Research in Higher Education, 49*(6), 513–530.

Taylor, J. C., Tseng, C. M., Murillo, A., Therrien, W., & Hand, B. (2018). Using argument-based science inquiry to improve science achievement for students with disabilities in inclusive classrooms. *Journal of Science Education for Students with Disabilities, 21*(1), 1–14.

The Iowa Assessments. (2015). *Research and development guide: Forms E and F*. Houghton Mifflin Harcourt.

Theriot, M. T., & Dupper, D. R. (2010). Student discipline problems and the transition from elementary to middle school. *Education and Urban Society, 42*(2), 205–222. doi:10.1177/0013124509349583

Tseng, C.-M. (2014). *The effects of the science writing heuristic (SWH) approach versus traditional instruction on yearly critical thinking gain scores in grade 5-8 classrooms.* (Unpublished doctoral dissertation). University of Iowa.

Woods, J. (2016). The fragility of argument. In F. Paglieri, L. Bonelli, & S. Felletti (Eds.), *The psychology of argument: Cognitive approaches to argumentation and persuasion* (pp. 99–127). College Publication.

Yackel, E., & Hanna, G. (2003). Reasoning and proof. In J. Kilpatrick, W. G. Martin, & D. Schifer (Eds.), *A research companion to principles and standards for school mathematics* (pp. 227–236). NCTM.

Yang, Y. T. C., & Chang, C. H. (2013). Empowering students through digital game authorship: Enhancing concentration, critical thinking, and academic achievement. *Computers & Education, 68*, 334–344.