Summer 2023

Undergraduates’ Preparedness for College-Level Work in STEM: The Importance of Reading and Understanding Scientific Theories, Arguments, and Data

Ewa M. Burchard
Claremont Graduate University

Undergraduates' Preparedness for College-Level Work in STEM:

The Importance of Reading and Understanding Scientific Theories, Arguments, and Data

By

Ewa M. Burchard

Claremont Graduate University

2023

Approval of the Dissertation Committee

This dissertation has been duly read, reviewed, and critiqued by the Committee listed below, which hereby approves the manuscript of the Ewa Burchard as fulfilling the scope and quality requirements for meriting the degree of Doctor of Philosophy in Education.

Dr. David E. Drew, Chair

Claremont Graduate University

Professor of Education

Joseph B. Platt Chair in the Management of Technology

Dr. DeLacy Ganley

Claremont Graduate University

Professor of Education

Dean of School of Educational Studies

Dr. June Hilton

Claremont Graduate University

Senior Research Fellow in Education

Abstract

Undergraduates' Preparedness for College-Level Work in STEM:

The Importance of Reading and Understanding Scientific Theories, Arguments, and Data

By

Ewa Burchard

Claremont Graduate University: 2023

This study focuses on undergraduates' preparation for college level courses. In recent international PISA results United States students fall behind seventeen countries on the computer-based reading proficiency test. They have scored low for over a decade, in spite of spending more than one hundred-thousand dollars per student on education. This US score is similar to, or lower than, the scores of other countries where spending is lower.

Considering reading performance statistics from the international PISA assessment and the inconclusive results from reading comprehension studies across media, the concern arose whether today's high school students are well prepared for college level courses in STEM. In response to these concerns, we assessed students' reading comprehension of a difficult scientific article. For that purpose, we investigated students understanding of text material on Earth's magnetism, across media. The assessment included reading a published article and assessment of comprehension, content knowledge and scientific argument quality. Over a hundred undergraduates in one-hundred-level Earth science classes responded to 29 multiple choice questions, in a regular class setting. Thereafter, reading, and scientific literacy strategies--comprehension, knowledge, content, and sourcing qualities--were evaluated.

Overall, participants performed similarly across media; however, several significant variations emerged between demographic groups. For example, females scored better than males on most strategies, and African American students outperformed Latinx students on most strategies. In general, students scored low on total understanding of the article, but higher on content knowledge than comprehension. Only 14 % of participants did well on understanding and argument quality, which coincided with higher familiarity with the topic (12% students) and higher interest (40% students), but the majority, overall, scored low on both familiarity and topic interest. Effective sourcing was correlated with high interest, understanding, and content knowledge.

The qualitative findings, from interviews, indicated, conversely, that students felt well prepared for college courses. They reported proficiency in both English reading and writing. Moreover, they reported enjoyment of taking college level courses.

This study suggests that it could be beneficial for students' entering college to become involved in peer reflection activities that would promote their scientific literacy learning, the skills gained when collaborating with others providing the opportunity for both scientific debates and self-reflection. Improvement of scientific literacy skills can increase the readiness of American students for college level work, including students from under-represented groups.

Dedication

I dedicate this dissertation to my dear mother.

Acknowledgment

I am tremendously grateful to Dr. David Drew, who believed in me from day one and created a safe space for me to discover researching at the doctoral level in my own learning style. Most importantly, he has been restlessly supported and guided me throughout my studies with great compassion and understanding.

I am grateful to Dr. June Hilton for her encouragement, responsiveness, and support during this research and the coursework.

I appreciate Dr. DeLacy Ganley, who generously accepted my invitation to serve as committee member and provided me with constructive feedback.

Much thanks also go to Dr. Alana Olschwang, Associate Vice President, of University Effectiveness, Planning, and Analytics, who generously took the time guiding me through the implementation of the data collection.

To my peers and colleagues, who accepted me to participate in their weekly video-discussion group, where we were sharing the progress and supporting each other, I owe a great debt of gratitude, especially Diana Edwards, Chia Her, Professor Steven Pell, and Denise Yaffe.

I want to thank Dr. John Keyantash, who opened up the opportunity for me to teach and meet a team of faculty members, mostly Jeff Cope and Cheyenne Cummings, as well as Dr. Gilamichel Domenico and Michelle Mohr, who were supportive of my survey and made a fundamental contribution to the data collection.

I am also grateful to my friend Nina Macaraig, who supported me in many ways and who reviewed and edited my documents.

Finally, I am grateful to the faculty at the CGU SES and TRNS; the four years of coursework at Claremont Graduate University prepared me well for conducting research.

Among them are Dr. Drew, Dr. Hilton, Dr. Adams, Dr. Pike, Dr. Maramba, Dr. Yamashiro, Dr. Luschei, Dr. Goode, Dr. Pedretti, Dr. Carter, Dr. Perkins, Dr. Camacho, and Dr. Bermudez.

Table of Contents

List of Figures

List of Tables

CHAPTER I Introduction

Statement of the Problem

This research contributes to global research on scientific literacy. The inspiration for such a study is derived from a decade of ongoing educational research emphasizing the importance of science comprehension and scientific communication, which are beneficial for future scientists as well as non-scientists. Developing scientific literacy in all students prepares them to serve society in the future, because these literacy skills contribute to an understanding of the quality of evidence and persuasive moves. Thus, they are equipped to make better judgements about political issues impacting them as voters, about the enactment of new policies, when forming decisions as jurors in court proceedings, or when electing government officials. This research shows that students continue to encounter challenges when selecting or formulating components that contribute to high-quality scientific argumentative writing.

Theoretical Rationale and Study Overview

This study evaluates students' argumentative skills in science, based on specific components representing levels of scientific argument quality, such as quality of the claim and quality of evidence. The quality of sourcing information is indicative of readers' abilities to make a "coherent mental representation" of the scientific text. According to Sandoval's Scientific Argument Model (Sampson and Clark, 2008), these components include: (1) content quality (types of evidence), and (2) quality of justification (excluded from the present study).

Although both are fundamental, they have been less studied constructs (Sandoval, 2010; Sampson and Clark, 2008).

Sandoval's Model, the *domain-specific model*, focuses on *field-dependent criteria*, which consist of *justification* and *content.* Since field-dependent criteria, as shared by the scientific community, include *conceptual and epistemological* qualities—which are *causal elements*, *complete explanations*, and *the data needed to warrant each element* (Sandoval, 2005; Sampson and Clark, 2008)—we have explored them with this assessment. Due to time and funding limitations, this research project has excluded the *justification* criterium, although it would warrant a follow-up investigation of the *types and levels of justification* of scientific argumentation.

In this research-study, the *content comprehension* components are based on Britt's Document Model (Britt et al., 1999), which include the ability of students to locate and assess the simplest constructs, such as *paraphrases* of a single idea, more complex *Intra-text Inferences* (source-content links), and most complex *Inter-text Inferences* (source-source links) (Britt et al., 1999; Salmeron, 2018). Per the *Construction-Integration Theory* (Kintsch, 1988; Britt et al., 1999; Salmeron, 2018), which constitutes a brick of Britt's Model, students create inferences while reading, and these are the *Coherent Mental Representation (CMR)* of phenomena within and across text(s). Thus, testing students' abilities to detect, locate, and evaluate these qualities constituted part of the survey. The questions asked students to evaluate levels of sourcing (from low to high): *general*, *specific*, and *embedded* (Britt, 1999; Salmeron, 2018). These two models served as tool for the assessment of students' comprehension, based on their own evaluation of scientific argumentative writing.

The idea for this study derived from both argumentative and systematic literature reviews on major topics: (1) the assessment of reading comprehension from various texts, (2) the assessment of scientific argumentative writing, and (3) the standardized assessment of students' performance in science, English, and math.

We entertained the idea of a relationship between independent variables representing students' demographics and reading conditions and the resulting variation in dependent variables representing their science comprehension and their own assessment of the quality of scientific argumentation.

In general, the comprehension variables included *single idea paraphrase*, *inferences*, and *synthesis*. These were fulfilled by looking at students' responses to *Intratext* and *Intertext Inferences*. A suggested follow-up study would also include *justification quality*, connections between claim and evidence (types), *persuading* with data (the quality of scientific evidence), as well as quality of *sourcing* (levels), as per Salmeron (2018). In accordance with the theoretical models, the components representing levels of comprehension and argument quality overlap with each other, such that content quality represents the quality of scientific argument, while based on previous studies sourcing and inferences directly relate to levels of comprehension.

The following variables were collected from the students' own evaluation of texts: claim, evidence, and sources. The analyses included descriptive and inferential statistics looking for mean variations under various conditions based on independent variables: demographics (race, gender, Pell Eligible, First Generation); students' perception of topic interest; topic knowledge; and paper/digital condition.

After assessing the scientific *content* and *sourcing (including inferences)* as dependent variables, the survey assessed the *argument quality (dependent variables)* following the guidelines of Sandoval's Scientific Argument Framework (Sampson and Clark, 2008) and Britt's Document Model (Britt, 1999; Salmeron, 2018).

Research Question

What is the level of undergraduates' preparedness for college when assessed on the basis of their comprehension of scientific texts under the conditions of students' background (demographics) and students' perception (topic familiarity and knowledge), with consideration of paper versus digital texts?

Study Significance and Contribution to Knowledge

Science education. This study is important for students who experience difficulty in comprehending scientific concepts (Salmeron, 2018; Sandoval, 2005; Sampson and Clark, 2008). Despite the increase of discourse practice in science classrooms, research shows that students continue to encounter problems when constructing high-quality scientific arguments. The demand for such a study is supported by ongoing research in science education and by the need for compliance with (relatively new) science education standards, the New Generation Science Standards (NGSS) published in the National Research Council (NRC, 2013), the benchmarks for science education. Every three years, the Programme for International Student Assessment (PISA, 1997) performs sets of high-stakes evaluations of reading and science literacy, by testing scholastic performance in mathematics, science, and reading. At the same time, the UNESCO Sustainable Development Goals (SDG 4, science education targets, 4.6 and 7) promotes inclusive, equitable literacy education in science and scientific communication, which includes scientific methods, ways of thinking, and understanding nuances in science (UNESCO).

Legal system. This research topic is also valuable for the legal system when it requires quality of evidence and data reliability in the context of criminal case proceedings. For example, some cases might rely on questionable scientific evidence, while others rely on witness testimonials. Therefore, the scientific community several years ago raised concerns about the quality of evidence that legal cases should use, while avoiding non-scientific evidence (Giannelli, 1993, p. 105). Koehler (2018) explains that scientific literacy is crucial for the legal system, because attorneys, prosecutors, jurors, and trial judges rely on it to arrive at decisions; Koehler suggests to reach out to the "scientific community for assistance when evaluating the sufficiency and reliability contributing to improvement of the proffered forensic methods that have long played an important role in the criminal justice system" (p. 29). Even in federal courts and some state courts, scientific reasoning plays an important role—for instance, Daubert's standards are used by attorneys, prosecutors, jurors, and trial judges who use scientific testimony by an expert witness, which constitutes just an opinion and not scientific evidence (Behr, n. d.). Reliable scientific evidence used in court include "DNA evidence, fingerprints, voice identification, bullet striations (markings), gunshot residue, hair and skin evidence, voice recognition, tire prints, and autopsy reports, which are all relevant to the finder of facts"; however, some of these examples are no longer used at court (for example, teeth marks from bites are no longer considered reliable evidence). Thus, greater scientific literacy assisting in recognizing reliable scientific evidence can save an innocent accused from an unfair guilty verdict.

Critical skills. Furthermore, as future generations become more and more exposed to digital reading, they need to develop the skills to quickly detect potential bias, flaws, or unreliability within digital texts (Breakstone, 2018). They need to be able to differentiate science from pseudoscience and recognize fraudulent or poorly done science before making decisions that might affect the well-being of society. Prior research has shown that the quality of scientific argumentation skills also aids in reading comprehension, enhances critical skills, and increases performance and engagement in academia.

Governmental policies, standards, and public safety. The above-mentioned examples show that science and reading literacy are required in science education as well as in other areas, such as the above-mentioned criminal justice system. Scientific literacy is emphasized on both global and national scales, in the Learning Framework of OECD 2030, the Sustainable Development Framework of UNESCO 2030, the Science Framework of K-12 in NGSS, and the Science Literacy Framework in PISA 2024. In addition, over twenty years of research have demonstrated the need for the development of proficient digital readers, due to the increased use of digital devices in academia, government, and industry. Expertise in digital reading is particularly crucial in times of national or international crises, during which immediate response is needed, as during pandemics or other natural disasters. Being able to recognize the quality and reliability of digital information and having a good science foundation can protect from disasters getting out of control. According to the National Oceanic and Atmospheric Administration (NOAA), success in protecting lives and properties depends on the ability to quickly detect and warn the government and the public (NOAA, 2016). In cases such as these, when reaction time is crucial, a proficient digital reader and scientifically literate person can quickly recognize the accuracy (OECD, 2021, p. 5), magnitude, and urgency of hazard warnings and make decisions on the most appropriate action.

Statistical data. The statistical data support the significance of this research, showing that in the United States only 12% (1/6) of adults scored at the highest level on standardized tests on basic literacy skills, which means that they have weak literacy skills (OECD, 2018, 2021). The OECD analyses report that, in middle- and low-income countries, the mean average science scores are concentrated below level 2, the baseline proficiency level as reported in PISA Brief, 2017 (OECD, 2017). At the same time, the use of digital devices for reading and internet searching has been exponentially rising; the 2012 study shows that the approximate time spent on devices increased from 21 hours per week in 2012 to 35 hours per week in 2018 (OECD, 2018, 2021).

Therefore, this research contributes to data concerning the impact of digital reading on comprehension, in response to a relatively recent implementation of educational reforms and new standards outlined in the Science Education Frameworks, NGSS, UNESCO 2030 Goals, and PISA 2024 Frameworks (OECD, 2020), aiming to improve students 'reading comprehension. It will also contribute to findings on the effect that the implementation of science literacy reform has on the freshman students 'outcomes in reading comprehension and argumentation. The results will inform us about the level of preparedness of underrepresented students for college-level courses, as the study aims to assess underrepresented undergraduates' skills acquired from pre-college education—that is, K-12 and social learning (OECD, 2018, 2021). Reading and science literacy research has gained more attention in education now.

The subsequent chapters are organized as follows: the literature review in Chapter II discusses global and regional perspectives on reading and scientific literacies, as well as on reading comprehension of multiple texts and scientific argumentative writing assessments.

Chapter III describes conceptual framework, target population, instruments, variables, research design, and the data gathering methods. Chapter III describes the mixed method research design, as guided by the conceptual framework developed in accordance with the reading and scientific literacies models that included a cluster sample data collection from the population of undergraduates in a four-year higher education institution. The instruments included the 29-question survey was prepared for the students in the lower-division Earth Science classes and a list of 11-interview questions. Next chapters, Chapter IV presents the findings, while Chapter V elaborates and concludes their meaning, it also lists limitations, and proposes the future study direction.

Glossary of Terms

AQSM: Argument Quality based on Sandoval's Model

AQFA: Argument Quality, Final Assessment

ACQ: Argument Content Quality

ASQ: Argument Source Quality

TAQ: Total Argument Quality

FAA: Final Argument Assessment

BDM: Britt's Document Model, sourcing-paraphrasing, inferring, and synthesizing

CIT: Construction Integration Theory—that is, creating inferences while reading

CMR: Coherent Mental Representation—that is, forming inferences within and between texts during reading

Conceptual Qualities: causal elements, presence and quality of claim, relevant and correct data; for example, if then, because, for, it follows, in respect to, according to, this is why…

CK: Content Knowledge

Daubert's Standards: use of expert witnesses' scientific testimony, an opinion

Epistemological Qualities: sufficiency of citations, presence of rhetorical references, holding causal coherence

EAR: Earth Science Class

Flesch-Szigriszt Index: text difficulty measure based on the INFLESZ scale (40-80)

First Generation: students who are the first in their family to attend college

Inter-text Inferencing: source-source link, synthesizing from multiple texts

Intra-text Inferences: source-content link, inferencing from one text

Justification: theory, model, concept, law, explanation

NGSS: New Generation Science Standards

NRC: National Research Council

NOAA: National Oceanic and Atmospheric Administration

OECD: Organization of Economic Cooperation and Development

Pell Eligible: undergraduates who display exceptional financial need

Pell Grant: a non-repayable federal subsidy awarded to students for post-secondary education, based on financial need

PISA 2024 Framework: plan to improve students' reading comprehension

PISA: Programme for International Student Assessment

RC: Reading Comprehension

SSAM: Sandoval's Scientific Argument Model, domain-specific with field-dependent criteria (justification, content)

SDG: Sustainable Development Goals

SQBM: Source Quality based on Britt's Model

STEM: Science, Technology, Engineering, and Math

Source Quality: types include general, specific, and embedded

Sourcing levels: paraphrasing, inferring, and synthesizing

Science Literacy: scholastic performance in math and science

Scientific Literacy: understanding nuances in science, scientific methods, and so on

TU: Total Understanding

UNESCO: United Nations Educational, Scientific and Cultural Organization

UNESCO 2030 GOALS: development of education for all

CHAPTER II Literature Review

Literature Roadmap

Despite the considerable growth of employment in computing and engineering, federal employment and education data continue to show an underrepresentation of Black and Hispanic professionals. According to the Pew Research Center (Pew, 2016-19, in Fry, Kennedy, and Funk, 2021), there is a long path ahead before the workforce in STEM diversifies. Statistical data show that Asian and white adults dominate STEM majors and workforce. In response to this structural imbalance, educators continue to search for tools that would increase underserved students' interest and success in the sciences. This study is grounded in the following argumentative and systematic literature review. The major topics investigated in scholarly publications include models of reading and scientific literacies, multiple texts' comprehension, scientific argumentations structures, underserved students' participation in STEM, and digital literacies.

The following body of literature has been identified for the organizational road map: (a) global to regional perspectives on reading and scientific literacies, (b) reading comprehension assessment, (c) assessment of scientific argument quality, and (d) students' literacy assessments.

Global to Regional Interest in Reading and Scientific Literacy

Improving science literacy has been of concern for over a decade. On a global scale, organizations emphasizing digital science literacy include UNESCO, with its Sustainable Development Goals 2030 (aiming for global access to education and an increase of digitalization), and OECD, which according to the OECD PISA 2024 Report (OECD, 2020) emphasizes the evidence-supported transition from paper and screen reading assessment to the exclusively digital assessment of multiple documents comprehension. On a national scale, these organizations are the National Science Teachers Association (NSTA), the American Association for the Advancement of Science (AAAS), the National Research Council (NRC), and Achieve, which have been developing standards with a focus on advancement in the sciences—for example, the New Generation Science Standards (NGSS) that have become an international benchmark for science education. The NGSS aims for important targets such as cross-cutting concepts, digital science literacy, and argumentation.

Reading Comprehension Assessments

According to the PISA 2024 Strategic Vision Direction in Science (SVDS) for youth and adults, science literacy is an important factor to assist in making decisions and actions to bring about change (PISA 2024 in OECD, 2020). Several internationally and nationally recognized science literacy frameworks have been put in place. The PISA Framework 2015 has incorporated three assessment components for reading from paper and screen, from low to high skills: (1) accessing and retrieving, (2) integrating and interpreting, and (3) reflecting and evaluating. In PISA 2018, the edited components included (I) single and (II) multiple digital texts comprehension. For (I) single text, sub-categories included (I-1) scanning and locating, (I-2) literal comprehension, (I-3) inferential comprehension, and (I-4) assessing quality and credibility reflecting on content and form. For (II) multiple texts, sub-categories subsumed (II-1) searching for and selecting the relevant text, (II-2) multiple text inferential comprehension, and (II-3) corroborating and handling conflicts (OECD, 2019, 2021). PISA 2018 no longer delineated paper or digital tasks (2019). Accordingly, the most important dimension of reading literacy is a cognitive process involving locating information, scanning a single text, and searching for and selecting relevant text from multiple documents (p. 39). Many studies have noted the value of prior content knowledge in the reading comprehension of various texts (Bartlett 1932; Nuttall, 1982; Bråten et al., 2011; Rouet, Britt, Mason, and Perfetti, 1996; Strømsø et al., 2010; Krammerer, 2019), whereas other studies (Salmerón, Gil, and Bråten, 2018) have posited that prior knowledge correlates to the text conditions. Over fifty studies, performed over a seventeen-year timeframe on ~170,000 participants, emphasized the benefit of reading on paper (Baron, 2016; Bucciarelli and Drew, 2015; Clinton, 2019; Kong, 2018; Aharony, 2018; Kazanci, 2015; Singer, 2018; Hermena, 2017; Mangen and van der Well, 2016; Dindar, 2017; Hoo, 2017; Margolin, 2013; Wolf, 2018). The PISA 2018 assessment results show the following: (1) strong readers perform well on both texts (OECD, 2018), (2) socio-economic status only shows on paper-based assessments (OECD, 2015), (3) digital reading marginally depends on navigation skills, (4) about 44% of variation in mean reading performance is positively associated with countries of higher per capita GDP (p. 32), (5) only 8.7% of students in OECD countries are top performers (Level 5-6); and (6) about 30% of association between socio-economic status and reading performance is accounted for by self-perception (OECD, 2019).

The gap in the performance on standardized tests between extreme groups (wealthy and poor) has been decreasing; however, as of now it is comparable to the three-year difference in schooling. These findings contradict the 2011 study projection, indicating that the gap is widening because of the income, and because the socio-economic status between poor and rich has been increasing (Reardon, 2011). The biggest gain is observed among middle school students who by the age of 17 equalize in their performance; no evidence of such gains has been recorded for students 17 and older (Hanushek et al., 2020, p. 3). These students will enter colleges and the future job market, requiring financial, social, cultural, and human capital resources (National Center for Education Statistics 2012a, in Hanushek et al., 2020, p. 3) and aspiring to opportunity (intergenerational mobility) for professional success (p. 3). Hanushek explains that the rise in inequality in the US and the widening of the income gap also trigger students' educational achievement gap (2020). According to Rothstein, income has "become more unequally distributed in the US in the last generation" (Rothstein, 2004, in Hanushek et al., 2020, p. 1). Historical trends in parental education, income, and ethnicity have been linked to students' achievement (Hanushek, 2020, p. 4), but not so much to the quality of schooling. Despite educational policy efforts, the gap in the achievement between children from high- and low-SES backgrounds has not changed. Thus, education researchers are concerned that, since cognitive skills remain critical for the economic well-being of US citizens, the gaps across socioeconomic measures do not project future improvements in intergenerational mobility. However, the 2022 report (Shakeel and Peterson, 2022; Shakeel, 2022) underlines that students of color achieve larger gains in math and reading scores on standardized tests than white students. This observation is based on research on 7 million students between 1971 and 2017 (Shakeel, 2022). The steepest increase among elementary school students of color, in particular, is in math (31% median performance gain). It has been noted that the "median rate of progress made by the average African American student is mostly noticeable in the elementary school students, but the gap diminishes in high school. African American students exceed white students by about 10 percent on the standard deviation per decade, in both reading and math" (Shakeel, 2022, p. 54).

These changes are explained by the growth in family income, parental education, and family size within the African American community, as well as "school desegregation, civil rights laws, early interventions like Head Start and other preschool programs, and compensatory education for low-income students" (A Half Century of Student Progress Nationwide, 2022). Latinx students' performance in math and reading shows an increasing trend when compared to white students, although these gains are less noted in reading than in math, due to the language barrier for Latinx students who account for 78 percent of English-language-learners in the US. Thus, this relatively recent report (Shakeel, 2022) concludes that the achievement gap based on socioeconomic status closes by 3% of a standard deviation per decade in both reading and math, and that the gaps in standardized assessments are narrowing between Latinx, African American, and Asian students who are making greater progress in reading and math than white students. In 2011, sociologist Sean Reardon raised alarm about the income-achievement gap; however, when comparing students' performance linked to their family income within a decade, between 2011 and 2022, this trend as not been confirmed. The evidence from 2022 compared across years show the qualitative similarity to the socioeconomic index (A Half Century of Student Progress Nationwide, 2022), which means that both are rising. This does not mean that the study does not show existing short-term discrepancies, because, among students who were tested at the end of high school, students from high-income families made greater progress in math and reading than those from low-income families (A Half Century of Student Progress Nationwide, 2022). However, there is a reverse relationship in the PISA study, which is assessing the application of math, science, and reading skills to solving real-life problems; it shows much larger gains for US high school students from families in the lowest socioeconomic status than for those in a higher economic status. In math, the performance of the most advantaged 15-year-old students decreased each decade by about 20%. Thus, there is a need for careful analysis of these results, which might be easily misinterpreted by the general public. The results need to be analyzed with specific attention to the type of tests (for example, PISA should not be compared to TIMSS), scale of diversity (for instance, percentage of immigrants), and the time-frame (long-term rate change versus immediate test results differ among participating nations). For example, according to the 2022 Report *A Half Century of Students' Progress Nationwide Report*, students in the bottom 25% of the socioeconomic distribution make more progress than their classmates in the top 25%, in both elementary and middle schools. In middle school, the difference is minimal between students, but by high school the highest socioeconomic group shows larger gains (A Half Century of Student Progress Nationwide, 2022). Also, when comparing the assessment results with family social status and income, the link was obvious, such that there was a direct relationship between the achievement and socioeconomic status; for example, parental educational attainment and family income were strong predictors of student achievement among students of different race (A Half Century of Student Progress Nationwide, 2022).

Shakeel (2022) explained that other aspects with the greatest impact on students' achievements consisted of school reforms, desegregation, accountability measures, more equitable financing, improved services for students learning English, and school choice.

Scientific Literacy Assessment - Scientific Argument Assessment

Salmeron (2018, p. 26) evaluated multiple documents on reading comprehension, based on students' memory of sources included in essays. Researchers, however, did not consider memory variations (such as sensory, short- and long-term), nor the quality of memories—such as haptic (touch), echoic (sound) and iconic (visual)—nor the variation in working memory—for instance, explicit (episodic, semantic) and implicit (procedural, associative, priming) (Camina and Güell, 2017). To avoid possible bias, this study focuses on reading comprehension as understanding and students' interaction with text, while activating sensory and short-term memory during writing. This study will not eliminate the impact of long-term memory, because long-term memory will be tested by using a questionnaire on topic knowledge or topic familiarity. Consistent with the document model framework (Britt et al., 1999; Perfetti et al., 1999; Rouet, 2006, and Strømsø et al., 2010, in Salmeron, 2018, p. 26), better comprehension is expected when reading from paper documents using haptic memory, as opposed to digital text (Salmeron, 2018) when haptic memory is not involved. While readers do not activate haptic experiences during digital reading, they might activate iconic and echoic memories. Historically, sourcing has been linked to understanding and the development of academic scholarship (Kopran et al., 1997; Norris, 1995, in Strømsø, 2013, p. 182). Strømsø (2018) explains that the reason why sourcing has received less attention in science learning (Yore et al., 2003) might be the fact that the epistemology of science is less related to interpretation and more to empirical testing (p. 183). Reliability, however, has been more central to scientific texts (p. 183).

The Source-Monitoring Framework (SMF), developed by Britt in 1999, represents comprehension as the formation of a mental map of a single document (intra-text model) and the mental map of multiple texts (inter-text model) assisting in detecting conflicting ideas (Salmeron, 2018, p. 26). Sandoval (Sampson and Clark, 2008), however, developed a scientific argument model, a discipline-specific model, which in addition to structure emphasizes justification and content. Thus, it is evaluated based on (1) epistemological quality, including (1a) sufficient amount of data, (1b) rhetorical references (inclusion, pointer, description, interpretation), and (1c) causal coherence; and (2) conceptual quality, which posits that (2a) the argument has a claim, (2b) causal elements are present in the claim, (2c) claim quality fits within the specific theoretical framework, (2d) data are present, and (2e) data are of high quality, or accepted by the scientific community (2008).

Toulmin's Argument Model (Kneupper, 1978) is the most frequently used discipline-general model. Toulmin's argument model focuses on structure, which is the presence of all components, such as claim, ground, warrant, qualifier, rebuttal, and backup. Another discipline-general model is Schwarz, Neuman, Gil, and Ilya's Model, which focuses on the structure and acceptability of reasons. Other types of models are domain-specific models, including Sandoval's, Zohar and Nemet's framework focusing on justification (as discussed in Sampson and Clark, 2008, p. 455), Kelly and Takao's framework focusing on the epistemic levels of propositions, and Lawson's framework, which focuses on the hypothetical-deductive validity of an argument (Sampson and Clark, 2008). Salmeron's framework was selected for this study because it emphasizes justification and content (such as data accuracy, validity, and reliability) and their relationships to claims; however, these qualities were less investigated in past studies, and the data contribute to the quality of the supporting evidence. Knight and Grymonpré's (2013) assessment of argument quality revealed four evidence groups: relevant/irrelevant, relevant-contradictory, high-quality scientific (data), and sufficient/non-sufficient. Additionally, Knight and Grymonpré (2013, p. 55) divided evidence into types related to socio-scientific context, such as scientific, ethical, political, and moral evidence (valued based on the argument's goal and audience). This is important because the scientific community places great value on empirical data (McNeill, 2014, p. 7). In other disciplines, justification can include empirical evidence, science ideas, appeals to authority, plausible mechanisms, and prior experience. The highest value (Level 1) of justification according to McNeill (2014) contains (1) empirical measurements and observations (scientific evidence) with supported, less valuable justifications, such as appeals to authority, or prior experiences. The middle level (Level 2) accounts for mixtures of justifications, which are combinations of more important forms of justifications (empirical measurements and observations) with less important forms (appeals to authority, prior experience), but they do not support, or just distract from, the flow of argument. Finally, the lowest level of justification (Level 3) contains less important justifications such as appeals to authority, personal stories, or plausible mechanisms (McNeill, 2014, p. 7). Another argument assessment can be found in the Berkeley Evaluation and Assessment Research (BEAR), a well-known Assessment System (Wilson, 2005, 2009, in McNeill et al., 2014, p. 5), which includes the following four steps: (1) construct maps (theoretical model of cognition), (2) items design (constructs in which theoretical construct is measured), (3) outcome space (possible ways in which students respond), and (4) measurement model (which translates item scores to the original construct map) (p. 5-6). In the BEAR assessment system, each building block is detailed for each of the five constructs: "(a) forms of justification, (b) relevant-supporting evidence, (c) sufficiency of the evidence, (d) multiple views, and (e) reasoning" (p. 6). Scientific argument is a logical and rational discourse, aimed at finding relationships between ideas and evidence (Duschl et al., 2007). It involves the development, evaluation, and validation of scientific knowledge (Driver, Newton, and Osborne, 2000), as well as knowledge construction (Ford, 2008). It comprises a claim (an explanation, conclusion), supported by evidence (data, observation), justified by rationale (principles, concepts, models). Students require the ability to understand and use (1) a conceptual framework (theories, principles, laws, models); (2) apply correct epistemology for evaluating a claim; and (3) construct and communicate knowledge as a social interaction process (Duschl, 2008). According to the NRC (2008), students lack the skills required for engaging in scientific argumentation. They struggle to find relevant data to support the claim and provide reliable evidence (McNeill and Krajcik, 2007). Knight and Grymonpré (2013) emphasize two forms of justification in the argument: *evidence* and *reasoning*, where the evidence consists of empirical data that support the claim, and reasoning uses scientific principles or ideas, while rebuttal critiques present the justifications for an alternative explanation (Knight and Grymonpré, 2013, p. 55). The scientific process involves the formation of (1) an argument with claims (evidence with data supporting the claim and justification); (2) an opposite argument; and (3) the scientist's counterargument and rebuttal of the skeptic's argument. Arguments are evaluated and critiqued based on the reliability and relevance of evidence, justifications using scientific explanations (such as scientific theory), or the quality of rebuttal of a skeptic's argument (validity of evidence, justifications, and explanations), and finally the quality of references—that is, their trustworthiness as scientific resources. Scientific argumentation can have a theoretical or practical purpose, where theoretical (classical) represents an oral or written disagreement, while practical serves the development of policies, tackling socio-scientific aspects (such as cross-cutting concepts). Argumentation in science education is important, since it increases motivation, content learning, debate skills, and knowledge-building practices (Erduran et al., 2005, in Telenius et al., 2020, p. 3). Science discourses help to present, support, refine, and evaluate scientific knowledge (Balland, 2016, in Telenius et al., 2020, p. 3).

Scientific evidence plays a crucial role in forensic ontology and in the fate of criminal cases. Kavitha et al. (2009) explain that evidence involved in legal proceedings needs to include scientifically reliable and sufficient data understood by jurors. These proceedings are based on expert testimonies, which might not provide scientifically reliable data; thus, the President's

Council of Advisors on Science and Technology (PCAST) recommended an action to strengthen forensic science and to promote its more rigorous use in the courtroom (Lander et al., 2016). In 2009, the National Academy of Sciences (NAS) released a report on forensic science; furthermore, in 2016 PCAST released a report, prepared by leading scientists, on the use of forensic science in criminal courts, calling for the use of reliable and sufficient scientific data (Koehler, 2008).

Educational researchers are concerned with students 'digital reading outcomes among diverse students of mixed reading skills and strategies, who struggle with or miss reading cues or controversial perspectives in digital documents (Salmeron, 2018). Prior research has shown significant discrepancies in reading proficiency between demographic groups of students (Wolf, 2018)—for instance, only 23% of Hispanic and 16% of African American students reach their reading potential, compared to 76% of students of European descent. Thus, this study aims to contribute to the available data by looking at those aspects that could increase underserved students' comprehension, competency, and relatedness to science.

Comparison of PISA and TIMSS Assessments

A large body of study exists on students' achievements on standardized tests, at both international and national scales, where scores from different tests in math, science, and literacy in participating countries were compared, without consideration of the countries' ratio of underrepresented groups, or the gaps between these groups. For example, the US with its diverse population of students, large income gap, immigrant groups with language barriers, and average score (American students scoring at average level) cannot be directly compared to countries that do not experience such diversity—such as China, with its students scoring first on each achievement test—until the social gap is resolved. Also, the standardized test presented below—although focusing on math, sciences, and reading—should be analyzed carefully, because, while TIMSS assesses basic math skills, PISA does not; rather, PISA assesses the application of math and reading to solving real-life problems. It is worth noting that PISA is not a basic math and reading test, but the measure of a 15-year-old person's preparedness for life. Contradictory to what TIMSS tests, PISA assesses students' ability to apply math skills in solving real-world problems; thus, 97% of PISA questions ask students to apply math to problems they might encounter in their everyday life, but only 48% such questions are included in the TIMSS assessment, which is rather focused on basic math skills (A Half Century of Students progress Nationwide, 2022). These tests provide apparent measures of student performance in OECD countries under specified criteria, but they should be used cautiously in rating educational achievements in each country and should not be directly compared with each other due to their varying circumstances. Another caution that needs to be considered while comparing is the short-term scores versus the rate of change in scores among groups—for example, in the US underprivileged students have historically driven down the mean score, especially since there is a large population of immigrants, half of which falls into lower socioeconomic groups. However, the rate of change in performance is greater among lower socioeconomic groups (African American, Latinx) than the rate of change among privileged students when looking at the longitudinal scale. Below are presented a few examples from previous research on student performance on these tests, including authors' concerns when it comes to students' achievement in respect to family social status, income, as well as students' gender, race, and grade level. These findings provide a good foundation for this study, which is concerned with students' preparedness for college.

In 2005, Boe and Shin, from the University in Pennsylvania, stated that "US students perform poorly on international comparisons of academic achievement" (Boe and Shin, 2005, p. 688). Similarly, Silver (1998) reported that seventh- and eighth-grade students performed poorly in mathematics and sciences on the TIMSS test (TIMSS, 1995 in Kelly, D. L., 2002). Thus, Rita Cowell, the director of the National Science Foundation, raised concern about the future of the US economy (Boe and Shin, 2005, p. 688).

This issue of American students not performing as expected has occurred for several years; however, international standardized tests, such as TIMSS and PISA, compare varieties of social groups that are not comparable since some have a much larger percentage of socioeconomically lower-status and racially more diverse populations than others. For example, in the US in 2009 there were about 19% immigrant students, and this population had grown to 23% by 2018. Among these immigrants, 40% were from poor families.

When considering the socioeconomic status among immigrant students with lower performance, the results decreased over time, but were worse in 2018 (OECD, 2019; PISA 2018). Although there was no significant difference in reading performance between immigrants and non-immigrants in the US, when accounting for the students' and schools' socioeconomic status, there was a difference, such that lower status was directly corelated with lower performance (OECD, 2019; PISA, 2018).

Boe and Shin's review of the results from international tests between 1991 and 2001 indicates that American students performed at about or above average in comparison to other industrialized countries, with a higher score in reading and civics and a lower score in math and an average score in science (Boe and Shin, 2005, p. 694). This is not as poor as it has been in the past, without considering the diverse population of immigrants and the large socioeconomic gap in the country. The lack of a full picture (Boe and Shin, 2005, p. 694), or choosing results that support the inadequacy of public education might have contributed to these alarming reports, for the purpose of promoting the privatization of public education (Boe and Shin, 2005). "Selection of supporting evidence is a common strategy in particular to advocate for a cause, policy and political direction," as Boe and Shin state (p. 694). Average performance is affected by racial and socioeconomic gaps (Boe and Shin, 2005), which has had a major impact and has now been constant for about a decade. In the US, the socioeconomically low group of immigrants, and Black and Hispanic students in particular, have been driving down the average for decades (Boe and Shin, 2005). Considering the statistics, in 1991 the population of public school students was about 70% white and 26% Black and Hispanic, but this ratio had increased by 2000, to 34% Black and Hispanic students, which is 8% during this 10-year period. Even if the achievement gap remained constant, the US mean scores could continue to decline as minority populations increase (Boe and Shin, 2005, p. 694).

The implementation of new education policies for raising student achievement on the national score may not have always brought beneficial outcomes, when compared at an international scale; for example, an earlier study by Lim and Sireci (2017) investigating the TIMSS, which is a basic skills test more so than PISA, focused on eighth-graders in the US showing improvement with respect to the "At or Above Basic" of the NAEP Achievement Level, but students in other countries improved more in the higher achievement level categories. This may be the outcome of the No Child Left Behind policy, which aimed to raise the low achievers to at least proficient level, but it did not improve the advanced mathematics achievement for eighth graders in the US, when compared to other countries (Lim and Sireci, 2017, p. 1). The TIMSS results from 2003, 2007, and 2011 suggested that the increase in mathematics achievement among US students was greater than the average increase for the other 31 countries with respect to the two lower achievement levels (Basic and Proficient). As Lim and Sireci state,

…

> The percentage increase of the U.S. students at or above the NAEP Proficient and Basic achievement levels were 5.9% and 5.3%, respectively, compared to 4.4% and 3.8%, on average, for the other countries. Regarding the NAEP Advanced level however, the percentage increase of the U.S. was smaller relative to that of the average percentage for other countries over the same period (2.9% for the U.S. versus the 4.6% average) (Lim and Sireci, 2017, p. 1).

Lim and Sireci's (2017) comparison of NAEP Basic and TIMSS Intermediate achievement levels showed similar increasing trends from 2003 to 2011, in the percentage of student At or Above both achievement levels. In regard to At or Above the NAEP Basic achievement level, the percentage of US students increased from 68.1% to 73.4%, while at or above the TIMSS intermediate level the percentage rose from 64.0% to 68.0%. When it comes to math, US students' average scores were at the NAEP basic level on the TIMSS mathematics tests in 2003, 2007, and 2011 (Lim and Sireci, 2017, p. 17).

According to the PISA reports, US teenagers' performance on international PISA tests has been stagnant since 2000. These results indicate that, among 15-years old students, 1/5 of the student read at the level of a 10-year-old. This is troubling news considering the fact that the labor market has heightened expectations for cognitive excellence from new graduates seeking employment.

These demographic trends have an enormous impact on students' achievement tests, and if the gap between underserved students and privileged students does not narrow, then US mean achievement scores will remain unchanged.

Students in the US perform at about or above the OECD average in reading and science, but below average in mathematics. US students' mean performance in reading has been stable since 2000, in mathematics since 2003, and in science since 2006, as it neither improved nor declined. The reading score, however, has shown a significant increase between 2009 and 2018 from about 4 to 13.5% (OECD, 2019; PISA 2018). Ethnic diversity and a gap in socioeconomic status places the average US student's score at a disadvantage when compared to less diverse countries (such as China).

In 2018, US students ranked 25th on average for all three subjects (PISA, 2018); more precisely, they ranked 37th in math, 18th in science, and 13th in reading. In comparison, Chinese students were reported as ranking 1st in each subject, but there is no report on income gaps among the Chinese population, nor the percentage of immigrants and underrepresented second language-speaking groups.

According to the PISA 2018 report, the performance gap between socioeconomically advantaged and disadvantaged students in reading was 99 score points (89 score points constitute the OECD average); about 27% of advantaged students in the US were top performers, but only 4% of disadvantaged students (OECD averages are 17% and 3%, respectively). However, when comparing the results within the US, only 10% of disadvantaged students were able to score in the top quarter (OECD, 2019; PISA 2018). In the US, socioeconomically advantaged students outperformed disadvantaged students in reading by 99 score points in the PISA 2018 study (with the OECD average being 89 score points). This is not significantly different from the average difference between the two groups (89 score points) across OECD countries. In the PISA 2009 report, the performance gap related to socioeconomic status was 107 score points in the US, and 87 score points on average across OECD countries.

There also exists a noticeable gender gap in reading performance, with female students outperforming male students, while in mathematics males did better than females, and in science both male and female students scored about the same. In terms of ratios, for every three male students only one female student plans to enter a profession in science and technology, such as engineering. In all participating PISA 2018 countries, girls significantly outperformed boys in reading, by 30 score points on average across OECD countries, while in the US the gender gap in reading (24 score points) was not significantly different from the average OECD gap. In the US, boys outperformed girls in mathematics by 9 score points. In comparison, across OECD countries, boys outperformed girls by 5 score points. While girls slightly outperformed boys in science (by 2 score points) on average across OECD countries according to PISA 2018, in the US girls and boys performed similarly (OECD, 2019; PISA, 2018).

> In the United States, 81% of students attained at least Level 2 proficiency in reading (OECD average: 77%). At a minimum, these students can identify the main idea in a text of moderate length, find information based on explicit, though sometimes complex criteria, and can reflect on the purpose and form of texts when explicitly directed to do so. (…) About 14% of students in the U.S. were top performers in reading, attained Level 5 or 6 in the PISA reading test (OECD average: 9%), but 73% of students in the U.S. attained Level 2 or higher in mathematics (OECD average: 76%). In the U.S. 8% of students scored at Level 5 or higher in mathematics (OECD average: 11%), which

indicates they can "model complex situations mathematically, and can select, compare and evaluate appropriate problem-solving strategies for dealing with them" (OECD, 2019; PISA 2018).

According to the OECD's 2019 report, the mean performance in reading, mathematics and science in the US has been about the same in every PISA assessment. Only the science performance in 2006 was below the 2018 mean score, but even in this subject performance has not changed significantly and has remained the same since 2009 (OECD, 2019; PISA 2018)

Promising results come from the 2018 PISA test, where there is a noticeable increase in performance: in reading, the group of 15-year-old students scoring at Level 5 or 6 (top performers) increased from about 4% (2009), to 13.5% (2018). In science, the lowest-achieving students also improved, narrowing the gap between lowest and highest performers. Equally, in science, the number of 15-year-old students scoring below Level 2 proficiency decreased by 5.7% between 2006 and 2018 (OECD, 2019; PISA 2018). This performance gap in the US is largely due to the country's diversity in terms of ethnicities and socioeconomic status, which then translates into a performance gap.

In a recent publication, Hanushek et al. (2020) stated that the socioeconomic status (SES) achievement gap has not been given much attention in prior research; thus, this gap is poorly documented (p. 1), and it is the reason why he compares US adolescent students' scores from the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Sciences Study (TIMSS), and the Program for International Student Assessment (PISA). Based on the SES index, Hanushek et al. (2020) measured the SES-achievement gaps between students from extreme groups (very poor and very rich), with the result that there are no changes (75-25 SES gap, SD=0.9) during a 50-year time period.

The gap between these extreme groups' performances is comparable to the three-year difference in schooling. These findings contradict Reardon's (2011) study, according to which the gap is widening. Apparently, the biggest gains were observed among middle school students, but no evidence has been recorded for students 17 and older (Hanushek et al., 2020, p. 3). Since students 17 and older are those who will be entering college and thereafter the job market, giving them financial, social, cultural, and human capital resources (National Center for Education Statistics, 2012a, in Hanushek et al., 2020, p. 3) and opportunities (intergenerational mobility) for professional success is paramount (p. 3). Hanushek et al. (2020) also explain that, as inequality in the US increases and the income gap widens, the gap in educational achievement by socio-economic status (SES) will also increase.

According to Rothstein, income has "become more unequally distributed in the United States in the last generation" (Rothstein, 2004, in Hanushek et al., 2020, p. 1). Historical trends in parental education, income, and race have been linked to students' achievement (Hanushek et al., 2020, p. 4), but not quite as much to the quality of schooling. Despite all policy efforts regarding education, the gap in achievement between children from high- and low-SES backgrounds has not changed. As long as cognitive skills remain critical for income and economic well-being of US citizens, the gap across the SES spectrum does not project future improvement in intergeneration mobility.

Interestingly, the 2022 report on students' progress noted larger gains in scores concerning performance on standardized tests in math and reading for students of color as opposed to white students (Shakeel, 2022). This conclusion is based on research studying 7 million students between 1971 and 2017. In particular, the biggest increase among students of color was observed in math among elementary school students (31% median performance gain).

It has been observed that the "median rate of progress made by the average black student is mostly noticeable in elementary school and diminishes in high school. African American student exceeds white student by about 10 percent of a standard deviation per decade in both reading and math" (Shakeel, 2022, n.p.).

These changes are explained by the educationally beneficial growth in family income, parental education, and family size among the African American community, as well as "school desegregation, civil rights laws, early interventions like Head Start and other preschool programs, and compensatory education for low-income students" (Shakeel, 2022, n.p.). Latinx students' performance in math and reading shows promising trends when compared to that of white students, although these gains are less noted in reading than in math, because the language barrier for Hispanic students contributes to about 78% of English-language-learners in the US. Thus, this recent 2022 report concluded that the achievement gap based on socioeconomic status closes by 3% of a standard deviation per decade, in both reading and math, and the gaps in standardized assessments are narrowing between Latinx, African American, and Asian students who are making greater progress in reading and math when compared to white students.

When comparing the 2022 and 2011 results linked to family income, the sociologist Sean Reardon raised alarm about the income-achievement gap, but this trend has not been confirmed, since evidence from 2022 points in the opposite direction, with the results showing qualitative similarity to the socioeconomic index (Shakeel, 2022). This does not mean that the study does not show existing short-term discrepancies because, among students who were tested at the end of high school, students from high-income families made greater progress in math and reading than those from low-income families (Shakeel, 2022). However, there exists a reverse relationship in the PISA study assessing the application of math, science, and reading skills, which shows gains for US high school students from families in the lowest socioeconomic status much larger than for those in a high economic status. In math, the performance of the most advantaged 15-year-old students has decreased each decade, by about 20 score points. Thus, there is a need for careful analysis of results, which might not be comprehensible to the general public, but to scholars who know to separate the scale (national versus international) and the time-frame (long-term rate of increase versus short-term results). For example, according to *A Half Century of Students' Progress Nationwide Report* (Shakeel, 2022), students in the bottom 25% of the socioeconomic distribution are making more progress than their classmates in the top 25% in both elementary and middle school. In middle school, the difference is minimal, but by high school the highest socioeconomic group makes larger gains. Also, when comparing the assessment results with family social status and income, the link was pronounced, such that there was a direct relationship between achievement and socioeconomic status—for example, parental educational attainment and family incomes were strong predictors of student achievement among students of different races.

Other aspects that have had the greatest impact on students' achievement were school reforms, desegregation, accountability measures, more equitable financing, improved services for students learning English, and school choice (Shakeel, 2022).

Literature Review Summary

This literature review has presented findings from global and regional perspectives on reading and scientific literacies, reading comprehension assessment and outcomes, scientific argument quality assessments, and a comparison of PISA and TIMSS literacy assessment results.

For over a decade, global and regional organizations concerned with education—such as UNESCO, OECD, AAAS, or NRC—have been emphasizing the progress and assessment of reading and scientific literacies as an effect of the technologically evolving world. According to the OECD, the PISA Strategic Vision of Direction in Science has been used as a tool for assessing students' development in science literacy, considered as a factor contributing to social and educational success. There is an ongoing interest in factors impacting students' science comprehension, as well as in assessment methods. Several testing approaches have been in place and keep evolving, such as locating information, integrating, interpreting, reflecting, and evaluating, either single or multiple ideas in one text or multiple texts. These approaches aim for detailed attention to key words in the texts, and while this practice is common in legal text reading, it has not been as obvious in scientific text reading; thus, students encounter challenges while analyzing and evaluating scientific argumentative writing, which, in fact, represent students' comprehension. In addition, many studies pointed to the value of prior knowledge and media, such as paper or digital texts; also, readers' background—such as gender, socioeconomic status, ethnicity, and self-perception—have been linked to students' performance, more so than schooling. The importance of such findings raises concern about graduates' future success, for cognitive excellence and scientific literacy are new qualities demanded by the technologically evolving job market. Considerable attention has been given to diversity of the workplace in

STEM, because there is a continued underrepresentation of people of color in multiple sciences and computer disciplines, in spite of their participation showing an increase much greater than the participation of white or Asian students.

Presentation of texts on paper and screen exposes students to different types of memory, since paper activates sensory memories or haptic experiences (touch, smell, visual), while the screen is only visual. Thus, reading from screen appears in the literature as intended less for enjoyment, and reading from paper is better memorized than digital reading. Students in many cases expressed their preference for reading a book from paper; however, while conducting a literature review of published articles, students preferred storing the literature digitally and reading from the screen, similar to students who read for math, science, and computer studies. Paper reading was preferred by older adults, while digital reading was the preference of the younger generation of readers. Students expressed the benefits of reading and in particular writing on the internet, such as social platforms and blogs. Overall, digital reading requires different skills than paper reading, and these strategies are being recognized and taught. Paper reading continues to be the preferred choice for reading novels for relaxation, while digital reading is more for work and study.

Reading comprehension has been associated with quality of writing; for example, a well-written scientific argument would indicate a high level of comprehension of scientific texts, which will then be reflected in the quality of claim, evidence, justification, and sources.

Apparently, a field-specific model of scientific argument is the best form of an argument to explain and move forward scientific research. Other, more general arguments do not fulfill that role so well. Thus, when evaluating scientific writing, one should apply the field-specific, scientific argument model, where a specific scientific theory or model justifies and connects the evidence with the claim. This approach requires students to grasp specific scientific theories, as well as their application and meaning; in effect, they are representative of students' high level of comprehension of scientific concepts. There are several quality levels of each component that have been used in evaluation or writing. The scientific argument is an important tool not only in science, but also in the legal system, where reliable evidence can save innocent peoples' lives. The results of standardized science literacy tests, perhaps not intentionally, may lead to misinterpreted findings; thus, researchers need to pay particular attention to the difference between individual students, scores, ratio of change, and the specifics of the assessment (basic skills vs applied).

The standardized test results of US students are lower than those of other, less diverse countries, such as China. In the US, about a quarter of the student population consists of students of color, often African American and Latinx students; these students score lower on standard literacy assessments, for their English-language skills are lower than the skills of students whose primary language is English. Thus, their performance lowers the overall national score. It is concerning that teenagers' performance has been stagnant for over 10 years, while job market expectations and demands for cognitive excellence have been growing with the development and use of technology in all areas. On one hand, the demand for new graduates increases, while on the other hand these graduates must fulfill the expectation of computer and STEM literacy. In effect, the topic of reading comprehension in the context of various media, scientific literacy, and student performance on assessments requires research investigating these components, as a reflection of students' preparedness for college-level classes. The following chapters will introduce the applied research design and methods.

Chapter III Research Design and Methodology

Overview

This mixed-method study combines a semi-experimental, exploratory quantitative approach with qualitative, semi-structured interviews in order to investigate students' preparedness for college-level courses, through an assessment of reading comprehension as reflected in the students' own evaluation of the quality of scientific argumentative writing. For that purpose, clusters from a population of undergraduates were drawn to complete a questionnaire after reading scientific texts in Earth Science classes. We received about 157 responses from students in 100-level undergraduate Earth Science classes (EAR 100 and EAR 101).

This study contributes to the literature on reading and scientific literacies by observing how variation in independent variables (IVs, such as students' background) causes changes in dependent variables (DVs, such as comprehension of texts). This allows drawing inferences about these relationships and assessing students' preparedness for college. We tested students' ability to evaluate a scientific article on Earth's magnetism and its arguments in order to assess whether their demographic characteristics can be linked to their performance.

The data set was analyzed through an independent t-test, by grouping independent variables—including (a) Gender, (b) Ethnicity, (c) First Generation, and (d) Pell Eligible—to perform analyses of significant differences in performance or links to dependent variables, such as reading comprehension (RC), content knowledge (CK), total understanding (TU), argument content quality (ACQ), final argument assessment (FAA), and sourcing (ASQ). Furthermore, students' interest in the topic (TI), familiarity with the topic (TF), and time exposure to digital reading (DRT) were also tested and served as additional independent variables (IV). This study shines light on the obstacles that students face when reading scientific texts and contributes to studies on comprehension—in effect, on their competency and relatedness to STEM education. The following section explains the conceptual models that assisted us in the quantitative assessment.

Conceptual Framework

Before deciding on the most suitable framework for this study, we considered several reading comprehension frameworks: (1) Schema Theory, which emphasizes the value of prior knowledge (Bartlett 1932; Nuttall, 1982); (2) Construction-Integration Theory (Kintsch, 1988), which emphasizes the role of semantics and syntax in the new text; (3) Sandoval's Domain-Specific Scientific Argument Model, which emphasizes justification and content (Sandoval and Millwood, 2005); (4) Britt's Document Model Framework, which focuses on the construction of a 'coherent mental representation' of ideas connected in a meaningful way, represented by intratext and intertext inferences (Britt et al., 1999; Perfetti, Rouet, and Britt, 1999; Rouet, 2006, in Salmeron, 2018); and (5) Johnson's Source-Monitoring Framework (Higgins and Johnson, 2012; Mitchell and Johnson, 2000), which concerns the encoding of effective, detailed, and distinct cues related to the document (cover, smell, touch, weight), characteristics contributing to increased memorialization, and "Text Annotation" and "Think Aloud" modes of text comprehension (which are similar but differ in data collection, since the former involves the coding of student's notes and the latter the coding of recorded comments). In effect, the components of Sandoval's and Britt's models greatly contributed to this study, which focuses on students' skills in the assessment of sources and scientific content; although the present study excludes justification, it proposes a follow-up survey on its role, because justification (for instance, in the context of scientific theories) is fundamental for the assessment of the level of comprehension of scientific concepts.

Prior studies identified seven reading strategies of highly effective readers, as listed below. Cognitive strategies—that is, mental processes used to extract and construct meaning from the text to create knowledge structures in the long-term memory—improve readers' comprehension and retention. A follow-up study could compare the average level of reading comprehension received from multiple assessments, the average quality of comprehension received from argument assessment, and the average level of comprehension from either "text annotation" or "think aloud," which constitute previously developed protocols with coding strategies. These include (a) activating, (b) inferring, (c) monitoring-clarifying, (d) questioning, (e) searching-selecting, (f) summarizing, and (g) visualizing-organizing.

The conceptual framework employed here is based on two models: (A) the Document Model Framework (Britt et al., 1999, in Salmeron, 2018), which guided the reading comprehension assessment based on sourcing, and (B) the Field-Specific Argument Quality Framework (Sandoval 2003, 2005, in Sampson and Clark, 2008), which guided the assessment of the quality of scientific argumentative writing.

The quantitative part of this study involved the collection of student responses about resurfacing phenomena in the texts, including authors, ideas, claims, documents (or sources, such as general sources, specific sources, and embedded sources), and understanding of the topic, including single-idea paraphrase, intra-text inferences, and intertext inferences. The results of the survey serve as data linking variations in students' reading comprehension and students' own assessment of the quality of scientific argumentative writing (scientific content and sourcing), in relation to the students' backgrounds. The qualitative section involved semi-structured interviews asking students how well-prepared they feel for college classes. Thus, the results from this research will benefit students, instructors, and evaluators.

With the help of these quantitative and qualitative frameworks, this study aims to answer overarching research questions, which concern students' preparedness for college.

Research Question

What is the level of undergraduates' preparedness for college when assessed on the basis of their comprehension of scientific texts under the conditions of students' background (demographics) and students' perception (topic familiarity and knowledge), with consideration of paper versus digital texts?

Methods

Sample and population of interest. The sample was collected from among the students taking lower-level undergraduate Earth Science courses from three different professors. The three professors were selected based on the number of 100-level courses that they taught in the Spring 2023 semester. The enrolled students who agreed to participate were assigned either paper or digital condition, such that EAR 101.04, EAR 101.09, EAR 101.08, EAR 101.01, and EAR 101.03 were assigned the digital text, resulting in 81 students (51.6%), while EAR 100.05 and EAR 100.03 were assigned the paper text, resulting in 76 (48.4%) of the whole sample N=157 students. Table 1 summarizes the demographic characteristic of the population at the research site located in Southern California, as retrieved from the institution's website (College Factual 2023).

Table 1

General Background Demographics of the Whole Population at the Research Site

Demographics	Number of students	Percent of a sample [%]
Males	6,660	35.6%
Females	12,027	64.36%
African American	2,018	10.80%
Latinx	11,956	63.98%
Asian	1,434	7.67%
White	1,322	7.07%
International	860	4.60%
Unknown	566	3.03%
Multiracial	454	2.43%
Pacific Islanders and American Indian	54	0.29%
Pell Eligible	10,950	58.6
First Generation	9,829	52.6%
Graduates	1,439	7.7%
Undergraduates	12,451	66.62%
Total	18,687 (2023)	100%

Note. Our sample of students was collected from undergraduate Earth Science classes in a four-year institution of higher education in Southern California. Sampled participants were 18 years and older. The age groups within the whole population range from 22-24 (3,802, 23.44%), 25-29 (3,768, 23,2%), < 18-19 (2,503, 15.43%), 20-21 (2,275, 14.03%), 35 and over (2,137, 13.18%), to 30-34 (1,511, 9.32%), as presented in the sample demographics in Chapter IV. Source: www.collegefactual.com (2023). Out of N=157 (100%) responders, about half of the students, 81 (51.6%), were surveyed under digital condition, and 76 (48.4%) under paper condition.

The study specifically aimed to include the underrepresented population of undergraduates; therefore, a four-year institution of higher education in Southern California was selected and approached. Once the institution agreed to be part of the study, the data collection was conducted in Spring 2023. The institution's overall population demographics (Table 3) is as follows: the total population size in 2023 is 18,687 students, out of which 12,451 are undergraduates and 1,439 graduate students. Among the undergraduates, 64.3% are women, and 35.7% men. In terms of ethnicity, 8,659 (69.5%) are Hispanic, 1,258 (10.1%) Black, 747 (6%) Asian, 603 (4.8%) International, and 585 (4.7%) white, while the remaining percentages are multiracial and unknown (College Factual 2023). About 90% of the student population at the institution are people of color, 64.0% Latinx, 10.6% African American, 7.9% Asian, 3.1% unknown, 6.6% white, 2.6% two or more races, and 3.4% Pacific Islanders and American Indian (CSUDH, 2020).

The following process in the sample selection was in place: (a) based on the research design aiming to test students on a topic in Earth Science, Earth Science classes were selected, and based on the researcher's interest in the student group at the lower college level, undergraduate courses were selected. Professors teaching several 100-level Earth Science classes were contacted. Three of the professors agreed to distribute the survey.

For the qualitative part of the study, students were personally invited to interviews in the classroom, by the professors and the researcher prior to or after the classes. An initial email was first sent to the professors who had already distributed the survey for the first quantitative section. Then the professors in their classes announced the ongoing interviews. In the following week, the researcher visited the classes and repeatedly made an additional announcement every day of the week. Some students were approached prior to the beginning of class or after the class had ended, and others while they were walking into or out of the science building. Prior to the interview, the students who agreed to participate in the interview first signed the consent form, then confirmed that they were at least 18 years old. All students who voluntarily answered survey questions were included in the data analyses and reporting.

The study's initial research design anticipated collecting data from at least 100 students taking undergraduate classes in Earth Science, and the survey received 157 responses (N=157). Participants took part in the reading sessions, either under paper condition (reading the article on paper) or under digital condition (reading the article from a screen). The conditions were assigned to specific classes; thus, students who were enrolled in lecture classes (EAR 100) received the paper article while completing the questionnaire online (Black Board), whereas students who were enrolled in the laboratory classes (EAR 101) received the article in digital format while also completing the questionnaire on Black Board. Before the assessment began, all students read and signed the consent form on paper as well as digitally. Hence, before they began the digital survey, all students had consented to anonymous data collection for further analyses and publication. All who agreed to be surveyed, N=157, completed and submitted their responses electronically. The number of responses fulfilled the proposed sample size. Thus, all responses in this study are from students who voluntarily agreed to participate. When compared to the college's population demographics, the student sample (N=157) is representative. Since undergraduate students are required to take lower-level science courses after transitioning from either High School or a two-year college, they were suitable for the assessment of science literacy and preparedness for college-level courses and future education.

Instruments and materials. Participants read an excerpt from a scientific article about the reversals of Earth's magnetism (K. Hoffman's 1988 article on "Ancient Magnetic Reversals: Clues to the Geodynamo," published in the journal *Scientific American*). The document was 596 words long (one single-spaced page). The excerpt did not contain any inaccurate or irrelevant information about the topic. The article was published in a reliable scientific journal; thus, according to the Flesch-Szigriszt Index and INFLESZ scale, it falls under the category of "very difficult" reading (Barrio-Cantalejo et al., 2008). In addition to the scientific article excerpt, the material consisted of the instrument (questionnaire) for the purpose of argument quality assessment with four shorter texts (366, 347, 300, and 315 words, respectively), prepared by the researcher, in the form of scientific arguments of various qualities.

These materials were prepared based on already existing measures: (a) standardized tests on reading comprehension and content knowledge, and (b) field-specific scientific argument and document source measures, which utilize the assessment of scientific argumentative writing and sourcing, respectively. These measures were combined to form the instrument for this study and to assess students' reading comprehension and content knowledge after reading the excerpt from the scientific article, as well as to assess the students' assessment of the quality of arguments after reading these scientific arguments. The final document included the article, three arguments, and 29 edited and new questions. The questionnaire was prepared based on previously validated instruments, albeit revised. Therefore, the questionnaire was also beneficial for the qualitative section, where students responded to a set of 11 questions on their perception of their preparedness for college. In the qualitative part of the study, a Zoom application was utilized for recording interviews on the computer, while transcribe.app assisted with transcribing the recorded narratives to a Microsoft Word document.

Assessment procedure. The study took place during Spring 2023 semester, after the Institutional Review Boards (IRB) at the two institutions notified the researcher that the proposed research protocol had been received, reviewed, and permitted, allowing the researcher to begin the survey. In accordance with the protocol, a half-hour time slot on one day per class was selected for reading the instrument and for digital data collection from the participants. Prior to the survey, students who agreed to participate in the study read and signed the consent form informing them about voluntary participation and the anonymous use of their responses as data in further analyses and publication. The consent form was provided to the students prior to the survey, in both paper and digital format. Following their agreement, the students began to read the materials and to respond to the digital questionnaire. The reading materials were presented to the groups in either digital or paper format. In the regular class setting, the questionnaire with 29 multiple-choice questions was accessible via the university website (Black Board), and a time limit of approximately half an hour was assigned. Thus, to complete the survey all students brought their digital devices. After completion of the survey, the professors exported the students' data to an Excel spreadsheet, formatted them by adding one column with Study Identifier, and then removed the participants' names. Thereafter, the students' responses were stored in a password-protected storage. Prior to the survey, the students were informed that they should take about 10 minutes to read the article excerpt and another 20 minutes to respond to the multiple-choice questions which included the text with the arguments (even though the article excerpt was also available to the students throughout). The instructions and the consent form, which also explained the purpose of the study, were included in the introductory section of the online questionnaire. The text was suitable for an undergraduate reading level.

An additional IRB review and approval took place for the qualitative section that was added later, and a new consent form was issued. Sixteen students signed the consent forms and responded to 11 open-ended questions (Appendix A).

Material distribution and process. An excerpt of 596 words was extracted from K. Hoffman (1988), "Ancient Magnetic Reversals: Clues to the Geodynamo," published in the journal *Scientific American*. Students received the text in a form that was enhanced by sources, such as in-text citations and two inscriptions (figures in Appendix A), to accommodate the purposes of this study. Furthermore, a question asking students to evaluate the quality of the scientific arguments required the addition of three scientific arguments of various lengths and qualities (with a word count of 331, 299, and 296, respectively; see Appendix A). The arguments with inscriptions contained one of five levels of evidence-sourcing (inclusion, pointer, description, assertion, or interpretations), according to the assessment of epistemological quality (Sandoval, 2005; Sampson and Clark, 2008), as detailed in the original version of the instrument containing divisions into specific analytical strategies and as explained also in Sandoval's model (2008).

The survey materials (digital article, printed article, and digital questionnaire) were distributed to the following 100-level Earth Science classes: EAR 100, two classes with approximately 50 students each; and EAR 101, six classes with approximately 24 students each. Students in EAR 100 received the paper article, while students in EAR 101 received the digital article, aiming for a distribution of 50% of students under each condition. While the initial number of students in the classes was 24 each for EAR 101 and 50 each for EAR 100, not all students participated; thus, we received 76 responses out of 100 from EAR 100 classes under the paper condition (N_p = 76), and 81 responses out of 144 from EAR 101 classes under the digital condition (N_d = 81), resulting in the sample total of 157 students (N=157).

The questionnaire first aimed at the assessment of Reading Comprehension (RC) and Content Knowledge (CK) after reading the article excerpt, then at the evaluation of the students' ability to identify the components of the field-specific scientific argument model (Sandoval, 2005; Sampson and Clark, 2008) in the article's arguments—such as the Argument Content Quality (ACQ), representative of the components' quality—and finally at identifying the components of the document sourcing model (Britt et al., 1999; Salmeron, 2018), such as the Argument Source Quality (ASQ), and performing the final argument assessment (FAA). At the end of the survey, students responded to the questions about Topic Interest (TI), Topic Familiarity (TF), and digital reading time (DRT) exposure. The instructions to students and professors are provided in Appendix A.

The interview questions aimed not only at assessing the baseline of students' preparedness for college, but also at learning about their strengths, major, college experience, and preferred choice of reading material.

Collection and types of data. This mixed-method study started with a semi-experimental, explanatory, and validation quantitative method to find the relationships between demographic groups' reading comprehension (including locating and evaluating the quality of the components of argumentative scientific writing). This was followed by a second, qualitative section with 11 semi-structured interview questions intended to explore the students' perception of their preparedness for college (Chapter IV). This approach required the researcher to collect data illustrating the effect that students' background has on the variation of their performance in science comprehension and on the students' own evaluation of the quality of scientific argumentative writing.

The following quantitative data sets were collected: Reading Comprehension (RC), Content Knowledge (CK), and Total Understanding (TU) as preliminary, followed by Argument

Content Quality (ACQ), Argument Sourcing Quality (ASQ), Total Argument Quality (TAQ), Final Argument Assessment (FAA), and Students' Background (Gender, Ethnicity, First Generation, Topic Interest [TI], Topic Familiarity [TF], and Digital Reading Time [DRT]). After collecting the students' responses to the multiple-choice questions, we received the variables coded as either correct or incorrect responses (1, 0) and the ordinal variables coded on the Likert Scale (see Table 2).

The following qualitative data were collected: students' narrative responses to interview questions about how their K-12 education prepared them for college classes, in terms of reading, writing, and science. Students' responses were then coded according to emerging themes and categorized into groups contributing to their motivation: competence, autonomy, and relatedness (see Chapter IV). The quantitative results from the statistical analyses were performed on the primary data derived from the assessment of students' reading comprehension and of the students' own evaluation of argument quality, as well as students' background collected from the secondary data source—that is, the institution's database. The following sections present the quantitative and qualitative results, starting with all variables, followed by descriptive statistics, inferential statistics, and finally the narrative qualitative study resulting from the semi-structured interviews at this institution of higher education with a predominant attendance of people of color.

Independent Variables. Independent variables were concerned with students' background—such as Gender, Ethnicity, Pell Eligibility, First Generation, and Parents' College—as well as primary variables, such as Students' Topic Interest (TI), Topic Familiarity (TF), and Digital Reading Time (DRT). Three student groups were defined: the whole sample, paper condition, and digital condition (categorized as a separate independent variable called Reading condition [P/D] used to split the sample into groups). Table 2 lists the variables and their designations.

Dependent Variables. The dependent variables included students' Reading Comprehension (RC), Content Knowledge (CK), and Total Understanding (TU), collected by means of questions related to scientific article comprehension and variables representing argument quality, such as Argument Content Quality (ACQ) and Argument Sourcing Quality (ASQ), the sum of which contributed to the Total Argument Quality (TAQ) variable. The data were collected through questions related to the evaluation of three scientific arguments and the final assessment of the scientific argument; this gave a separate data set called Final Argument Assessment (FAA). Table 2 lists the variables and their designations.

The analyses were based on the following sets for combined dependent variables: (a) RC, CK, and TU for the whole sample, without separation into groups; (b) RC, CK, and UT, separated by Gender; (c) RC, CK, and TU, separated by Ethnicity group (African American versus Latinx); (d) RC, CK, and UT, separated by First Generation (FG) students versus students whose parents attended or graduated from college; and (e) RC, CK, and TU, separated by Pell Eligible groups.

Reading Comprehension (RC). Reading Comprehension consisted of four questions (Q1-Q4) asking about persuasive moves (strengthening and weakening, Q1, Q3), inferences (Q4, Q5), and claim and/or conclusion (Q2) in the primary reading material—that is, the excerpt from the scientific article provided to the students (Appendix A).

Content Knowledge (CK). Content knowledge was comprised of a combined twelve variables based on four questions (Q5-Q16) asking about the scientific content (Q5-Q12), as well as about the content as evidence in support of a claim (Q13-Q16) directed towards the primary reading text. While questions Q5-Q12 were concerned with the scientific content, questions Q13-16 were concerned with the role that the scientific content played in support of the claim and evidence represented by Argument Content Quality (ACQ) and Argument Sourcing Quality (ASQ) for the argument analysis. However, because it related to the primary reading excerpted from the article, it was included together with Q5-12, which were related to the Content Knowledge (CK) variables.

Total Understanding (TU). The variable representing the students' understanding of the text is a combined variable of both RC and CK. The TU variable, as well as RC and CK, were further compared to the four-scale grading system (approaching, proficient, emerging, advanced) to better explain students' performance and to report the findings.

Table 2

Variable Groups

Variable Names and Designation	Coding
Reading Comprehension (RC): Q1-4	0-4
Content Knowledge (CK): Q5-16	0-12
Total Understanding (TU): Q1-Q16	0-16
Argument Content Quality (ACQ): Q17-20	0-4
Argument Source Quality (ASQ): Q21-25	0-5
Total Argument Quality (TAQ): Q17-Q25	0-9
Topic Interest (TI): Q26 (Likert: not interesting, somewhat, interesting, very interesting)	0,1,2,3

Topic Familiarity (TF): Q27 (Likert: unfamiliar, somewhat, familiar, very familiar)	0.1.2.3
Digital Reading Time (DRT): Q28 (<5, 5-10, >10, and empty >20 YRS)	0,1,2,3
Final Argument Assessment (FAA): Q29	1,2
Gender (G)	1,2
Ethnicity (E)	1,2
Pell Eligible (PE)	0,1
First Generation (FG) and Parent College (PC)	1,2
Paper Digital (PD)	1,2

Note. RC - Reading Comprehension was coded as scale 0-3: approaching (0), proficient (1), emerging (2), advanced (3); CK - Content Knowledge was coded as scale 0-11: approaching (0-1), proficient (2-4), emerging (5-6), advanced (7-11); TU - Total Understanding was coded as scale 0-16: approaching (0-1), proficient (2-5), emerging (7-9), advanced (10-16); ACQ - Argument Content Quality was coded as scale 0-5: very low (0), low (1), medium (2), high (3), very high (4), exceptional (5); ASQ - Argument Source Quality was coded as scale 0-5: very low (0), low (1), medium (2), high (3), very high (4), exceptional (5); TAQ - Total Argument Quality was coded as scale 0-5: very low (0), low (1), medium (2), high (3), very high (4), exceptional (5); FAA - Final Argument was coded as dichotomous variable; G - Gender, E – Ethnicity (E), PE - Pell Eligible, FG - First Generation, PC - Parent College, PD - Paper/Digital Condition

CHAPTER IV Results

Quantitative Results

This section will present the descriptive statistics concerning the background variables and key variables of the analyses. It will then present the inferential statistical analyses of the data and conclude with the results from the series of interviews conducted with the students.

Descriptive Statistics

Analytical results of the three strategies factored by student groups. The results give a better understanding of the differences in students' performance on RC, CK, and TU (dependent variables) within groups of various background characteristics, such as Gender, First Generation, Ethnicity, and Pell Eligible (independent variables). The following section explains these findings.

Descriptive results for demographics. The sample of N=157 students attending undergraduate Earth Science classes at a four-year public institution of higher education was analyzed to obtain descriptive and inferential statistics. Table 3 summarizes the demographic data obtained from the descriptive analyses. According to these results, the group of surveyed students reflects the characteristics of the population in this public institution of Higher Education in Southern California. The predominant ethnicity consists of students of color; in the sample, there were included 120 (76.4%) Latinx, 21 (13.4%) African Americans, 5 (3.2%) Asian, 3 (1.9%), White, 3 (1.9%) multi-race or unknown, and 3 (2.1%) non-resident students (Table 3).

The sample data identified that the majority consisted of First Generation students (92, 58.6%); less than one quarter of the students had parents with some college (29,18.5%), and even fewer students had parents who graduated from college (17, 10.8%). In the sample, about half of the students were Pell Eligible (92, 58.6%), which classifies them as having low socioeconomic status, while the remaining students in the sample (58, 36.9%) were non-Pell Eligible; thus, they are classified as medium to high socioeconomic status (Table 3).

Our sample of undergraduates from Earth Science classes at a four-year institution of higher education in Southern California, with a total population of 18,687 (2023), as seen in Table 3, reflects the characteristics of the university population. The sampled participants were eighteen years and older. The age groups of the entire population, listed by size, ranges as follows: 22-24 (3,802, 23.44%), 25-29 (3,768, 23,2%), 18-19 and under (2,503, 15.43%), 20-21 (2,275, 14.03%), 35 and over (2,137, 13.18%), and 30-34 (1,511, 9.32%) (Collegefactual.com, 2023).

Table 3

Demographics of the Sample (N=157) versus population (N'=18,687) at the research site

Demographics	Number of students	Percent in sample N=157 [%]	Population statistics Based on N'=18,687 students [%]
Male	52	33.1	35.6
Female	104	66.2	64.3
African American	21	13.5	10.8
Latinx	120	76.4	64.0
Asian	5	3.2	7.7
White	3	1.9	7.1
International	4	2.5	4.6
Multiracial	3	1.9	2.4
Pell Eligible	92	58.6	58.6

Demographics	Number of students	Percent in sample N=157 [%]	Population statistics Based on N'=18,687 students [%]
First Generation	92	58.6	52.6
Parents some college	29	18.5	-
Parents graduated from college	17	10.8	-

Note. N'=population, N-sample

Descriptive results on total understanding and argument quality in the whole sample. The descriptive statistics included results from the assessment of comprehension of the scientific article and the scientific arguments analyses. The following analyses of reading comprehension, content knowledge, total understanding, and argumentative quality were obtained:

Reading Comprehension (RC). The descriptive statistics for the entire sample N=157 for RC, with the maximum possible score of 4, has the mean value 1.24 and SD .802. This indicates that students' reading comprehension is low in the entire sample, while the greater number of scores concentrate around level 2 of RC proficiency, given that level 1 was coded as 0, level 2 as 1, level 3 as 2, and level 4 as 3, based on a four-level grading scale (approaching, proficient, emerging, and advanced). This places half of the students from the sample at a low level of comprehension. The distribution among estimated proficiency levels was as follows: level 1-approaching (0) N=29 (18.5%), level 2-proficient (1) N=69 (43%), level 3-emerging (2) N=52 (33.1%), and level 4-advanced N=7 (4.5%) (Table 4).

Content Knowledge (CK). The descriptive statistics for the entire sample N=157 for CK, with the maximum possible score of 12, shows that the mean value for the sample is at 7.69 and SD 1,724, indicating that about half of students on average scored at levels 7-8. On the same four-level scale for content knowledge (very low 3, low 6, medium 9, and high 12), this would place 50% of the students at the level low to barely medium, 20% at the very low to low level, 23% at the medium level, and only 2% at the high level, when it comes to the content knowledge test presented in the survey in Table 11. Content knowledge (with 0-11 possible points) was assigned four levels, as follows: level 1-approaching (0,1), level 2-proficient (2-4), level 3-emerging (5-6), and level 4-advanced (7-11), corresponding to the four-level scale listed above.

Table 4

Descriptive statistics for combined variables representing total understanding

Variable	N	Mean	Std. Deviation
Reading Comprehension (RC)	157	1.24	.802
Content Knowledge (CK)	157	7.69	1.724
Total Understanding (TU)	157	8.92	1.966
Valid N (listwise)	157		

Note. Reading Comprehension is the sum of Q1-Q4_RC, Content Knowledge is the sum of Q5-Q16_CK, while Total Understanding is the sum of Reading Comprehension and Content Knowledge. Whole sample.

*Total understanding (TU)***.** The descriptive statistics for Total Understanding (TU), which is comprised of RC and CK, resulted in the following findings for the whole sample N=157, with the maximum possible score of 16: the mean value for the sample was at 8.92, and SD 1.966. This indicates that 50% of students scored at levels 8-9 on a four-level scale (very low 4, low 8, medium 12, and high 16); thus, half of the sample falls in the low level on the measure of Total Understanding (Table 4). TU (with a possible 0-16 points) was assigned to four levels, as follows: level 1-approaching (0,1), level 2-proficient (2-6), level 3-emerging (7-9), and level 4-advanced (10-16), corresponding to the four-level scale listed above. In the entire sample (N=157), the maximum score for TU was 14, and the minimum 4.

The following tables present descriptive statistics for Reading Comprehension, Content knowledge, Total Understanding, Argument Content Quality, Argument Sourcing Quality, Total Argument Quality, Final Argument Assessment, Topic Interest, Topic Familiarity, and Digital Reading Time, grouped into students' demographic groups and organized in the following order: Gender, Ethnicity, First Generation, and Pell Eligibility.

Descriptive results on total understanding and argument quality in gender groups. The descriptive statistics revealed that, when considering gender, the Total Understanding (TU) mean value was higher for females than for males; findings were similar for reading comprehension (RC), where female scored higher than male students. However, in terms of the content knowledge (CK), male performed better than female students (Table 5).

Table 5

Descriptive statistics for Total Understanding, Whole Sample by Gender

Variable	Gender	N	Mean	Std. Deviation
Reading Comprehension	Female	104	1.27	.850
	Male	52	1.15	.697
Content Knowledge	Female	104	7.68	1.742
	Male	52	7.71	1.719
Total Understanding	Female	104	8.95	2.036
	Male	52	8.87	1.858

Notes. Reading Comprehension is the sum of Q1-Q4_RC, Content Knowledge is the sum of Q5-Q16, while Total Understanding (TU) is the sum of RC and CK.

On Argument Quality, such as ACQ, TAQ, and FAA, female scored higher than male students, except for ASQ, on which male scored higher than female students (Table 6).

Table 6

Descriptive Statistics for Argument Quality and Assessment, Whole Sample by Gender

Variable	Gender (G)	N	Mean	Std. Deviation
Argument Content Quality (ACQ)	Female	103	2.29	.812
	Male	52	2.00	.767
Argument Sourcing Quality (ASQ)	Female	103	1.03	1.071
	Male	52	1.12	.963
Total Argument Quality (TAQ)	Female	103	3.32	1.315
	Male	52	3.12	1.263
Final Argument Assessment (FAA)	Female	102	.49	.502
	Male	52	.33	.474

Notes. ACQ – Argument Content Quality, ASQ – Argument Source Quality, TAQ – Total Argument Quality, FAA – Final Argument Quality

Other variables included students' perception of their interest in the topic and familiarity with the topic, as well as how long they had been exposed to digital reading. According to these data, male students were more interested in and more familiar with the topic, and they also reported reading digitally for a longer period of time than female students (Table 7).

Table 7

Descriptive Statistics for Topic Interest, Familiarity and Digital Reading Time, Whole Sample by Gender

Variable	Gender (G)	N	Mean	Std. Deviation
Topic Interest (TI)	Female	102	1.32	.798
	Male	52	1.46	.939
Topic Familiarity (TF)	Female	102	.57	.764
	Male	51	.69	.883
Digital Reading Time (DRT)	Female	102	.70	.672
	Male	51	.92	.771

Descriptive results in ethnicity groups. Based on the ethnicity of students and their scores on dependent variables RC, CK, TU, ACQ, ASQ, TAQ, FAA, TI, TF, and DRT, we have learned that African American generally scored higher than Latinx students. African American students scored higher on TU, RC, and CK. These two ethnic groups account for approximately 95% of the entire sample; thus, their performance results were of more interest than the remaining 5% of the other ethnicities in the sample (Table 8).

Table 8

Descriptive Statistics for Total Understanding, Whole Sample by Ethnicity

Variable	Ethnicity	N	Mean	Std. Deviation
Reading Comprehension	African American	21	1.29	.717
	Latinx	120	1.19	.792
Content Knowledge	African American	21	7.90	1.758
	Latinx	120	7.67	1.751
Total Understanding	African American	21	9.19	2.015
	Latinx	120	8.86	1.980

Notes. Reading Comprehension is the sum of Q1-Q4_RC, Content Knowledge is the sum of Q5-Q16, while Total Understanding (TU) is the sum of RC and CK.

African American students scored higher than Latinx on ASQ and TAQ, while Latinx scored higher on ACQ and FAA (Table 9).

Table 9

Descriptive Statistics for Argument Quality and Assessment, Whole Sample by Ethnicity

Variable	Ethnicity (E)	N	Mean	Std. Deviation
Argument Content Quality (ACQ)	African American	20	2.00	1.026
	Latinx	120	2.18	.788
Argument Sourcing Quality (ASQ)	African American	20	1.30	.865
	Latinx	120	1.03	1.028
Total Argument Quality (TAQ)	African American	20	3.30	1.380
	Latinx	120	3.22	1.304
Final Argument Assessment (FAA)	African American	20	.35	.489
	Latinx	120	.44	.499

Notes. ACQ - Argument Content Quality, ASQ - Argument Source Quality, TAQ - Total Argument Quality, FAA - Final Argument Quality.

African American scored higher than Latinx students in TF and DRT, while Latinx students scored higher in TI.

Table 10

Descriptive Statistics for Topic Interest, Familiarity, and Digital Reading Time, Whole Sample by Ethnicity

Variable	Ethnicity (E)	N	Mean	Std. Deviation
Topic Interest (TI) Q26	African American	20	1.25	.639
	Latinx	120	1.39	.873
Topic Familiarity (TF) Q27	African American	20	.75	.910
	Latinx	119	.62	.802
Digital Reading Time (DRT) Q28	African American	20	.95	.686
	Latinx	119	.71	.717

Notes. TI -topic interest, question Q26, TF-Topic Familiarity, Q27, and DRT- Digital Reading Time Q28.

Descriptive results in First Generation and non-First Generation students. TU and RC were lower for First Generation students than for non-First Generation students (students whose parents had some college and/or graduated from college); however, on CK, First Generation students scored higher than non-First Generation students.

Table 11

Descriptive Statistics for Total Understanding, Whole Sample by First Generation vs Parents some College and Graduated from College

Variable	First Gen vs Parent College & Graduated	N	Mean	Std. Deviation
Reading Comprehension	First Generation	92	1.13	.801
	Parent College/Grad	51	1.41	.853
Content Knowledge	First Generation	92	7.84	1.817
	Parent College/Grad	51	7.71	1.514
Total Understanding	First Generation	92	8.97	2.019
	Parent College/Grad	51	9.12	1.872

Notes. Reading Comprehension is the sum of Q1-Q4_RC, Content Knowledge is the sum of Q5-Q16, while Total Understanding (TU) is the sum of RC and CK.

First Generation students performed better on argument quality assessment, such as on ASQ and TAQ, but they scored lower on ACQ and FAA (Table 12).

Table 12

Descriptive Statistics for Argument Quality, Whole Sample by First Generation vs Parents some College and Graduated from College

Variable	First Generation /Parent College w/Graduated	N	Mean	Std. Deviation
Argument Content Quality (ACQ)	First Generation	91	2.21	.823
	Parent College/Grad	51	2.24	.764
Argument Sourcing Quality (ASQ)	First Generation	91	1.15	1.105
	Parent College/Grad	51	.92	.977
Total Argument Quality (TAQ)	First Generation	91	3.36	1.426
	Parent College/Grad	51	3.16	1.084
Final Argument Assessment (FAA)	First Generation	91	.40	.492
	Parent College/Grad	51	.47	.504

Notes. Argument Content Quality, ASQ - Argument Source Quality, TAQ - Total Argument Quality, FAA - Final Argument Quality, FG-First Generation vs parent some college/parent graduated college, which includes parent some college and/or parent graduated from college.

First Generation students also reported higher interest in the topic and were more familiar with the topic; moreover, they had also experienced longer time exposure to digital reading (Table 13).

Table 13

Descriptive Statistics for Topic Interest, Familiarity and Digital Reading Time, Whole Sample by First Generation vs Parents some College and Graduated from College

Variable	First Generation /Parent College w/Graduated (FG1)	N	Mean	Std. Deviation
Topic Interest (TI) Q26	First Generation	91	1.43	.871
	Parent College/Grad	51	1.29	.832
Topic Familiarity (TF) Q27	First Generation	90	.64	.852
	Parent College/Grad	51	.59	.779
Digital Reading Time (DRT) Q28	First Generation	90	.78	.761
	Parent College/Grad	51	.80	.633

Notes. Argument Content Quality, ASQ - Argument Source Quality, TAQ - Total Argument Quality, FAA - Final Argument Quality, FG - First Generation vs parent some college/parent graduated college; Parent college includes parent some college and/or parent graduated from college.

Descriptive results in Pell Eligible and non-Pell Eligible students. When considering Pell Eligibility, TU and RC were lower for Pell Eligible students, but students in this group scored higher on CK (Table 14).

Table 14

Descriptive Statistics for Total Understanding, Whole Sample by Pell Eligible

Variable	Pell Eligible	N	Mean	Std. Deviation
Reading Comprehension	Pell Eligible	92	1.13	.730
	Non-Pell Eligible	65	1.38	.878
Content Knowledge	Pell Eligible	92	7.78	1.602
	Non-Pell Eligible	65	7.55	1.888
Total Understanding	Pell Eligible	92	8.91	1.801
	Non-Pell Eligible	65	8.94	2.193

Notes. Reading Comprehension is the sum of Q1-Q4_RC, Content Knowledge is the sum of Q5-Q16, while Total Understanding (TU) is the sum of RC and CK.

Furthermore, when considering Pell Eligibility, ACQ, TAQ, and FAA scores were lower for Pell Eligible students, but this student group scored higher on ASQ (Table 15).

Table 15

Descriptive Statistics for Argument Quality and Assessment, Whole Sample by Pell Eligible

Variable	Pell Eligible (PE)	N	Mean	Std. Deviation
Argument Content Quality (ACQ)	Pell Eligible	91	2.11	.836
	Non-Pell Eligible	58	2.36	.742
Argument Sourcing Quality (ASQ)	Pell Eligible	91	1.11	1.069
	Non-Pell Eligible	58	.93	.814
Total Argument Quality (TAQ)	Pell Eligible	91	3.22	1.389
	Non-Pell Eligible	58	3.29	1.060
Final Argument Assessment (FAA)	Pell Eligible	91	.41	.494
	Non-Pell Eligible	57	.46	.503

Notes. ACQ - Argument Content Quality, ASQ - Argument Source Quality, TAQ - Total Argument Quality, FAA - Final Argument Assessment

Pell Eligible students also reported higher familiarity with the topic, but lower interest in the topic and a shorter time exposure to digital reading (Table 16).

Table 16

Descriptive Statistics for Topic Interest, Familiarity and Digital Reading Time, Whole Sample by Pell Eligible

Variable	Pell Eligible (PE)	N	Mean	Std. Deviation
Topic Interest (TI) Q26	Pell Eligible	91	1.36	.863
	Non-Pell Eligible	64	1.39	.828
Topic Familiarity (TF) Q27	Pell Eligible	90	.68	.805
	Non-Pell Eligible	64	.52	.797
Digital Reading Time (DRT) Q28	Pell Eligible	90	.72	.735
	Non-Pell Eligible	64	.84	.672

Notes. TI-Q26, TF-Q27, DRT-Q28

RC, CK, and TU mean results comparison among groups. The descriptive statistics informed by studying the whole sample N=157 of undergraduate students, for the dependent variables TU (total understanding), RC (reading comprehension), CK (content knowledge), under consideration of the independent variables G (Gender), E (Ethnicity), FG (First Generation), and PE (Pell Eligibility), are as follows:

Total Understanding among demographic groups. The TU mean value, comprised of RC and CK, was 8.92 out of a maximum score of 16, which on the four-level grading scale (approaching, proficient, emerging, advanced) falls on level 3 (emerging); female students scored higher than male students (8.95 vs 8.97), African American students scored higher than Latinx students (9.19 vs 8.86), First Generation students scored lower than non-First Generation students (8.97 vs 9.12), and Pell Eligible students scored lower than non-Pell Eligible students (8.91 vs 8.94).

Reading Comprehension among demographic groups. The RC mean value was 1.24 out of the maximum score of 3, which indicates that students in the sample fall on the proficient level (low level). Female scored higher than male students (1.27 vs 1.15), African American scored higher than Latinx students (1.29 vs 1.15), First Generation students scored lower than non-First Generation students (1.13 vs 1.41), and Pell Eligible students scored lower than non-Pell Eligible students (1.13 vs 1.38), while the digital condition presented 1.20 vs 1.28 in the paper condition.

Content Knowledge among demographic groups. The CK mean value for the sample was 7.69 out of a maximum of 12, which indicates that students in the sample fall on the advanced level. Female scored higher than male students (7.68 vs 7.71), African American scored higher than Latinx students (7.90 vs 7.67), First Generation students scored higher than non-First Generation students (7.84 vs 7.71), and Pell Eligible students scored higher than non-Pell Eligible students (7.78 vs 7.55).

ACQ, ASQ, TAQ, and FAA mean results comparison among groups. Based on the descriptive and inferential statistics of the whole sample N=157 of undergraduate students, for the dependent variables ACQ (argument content quality), ASQ (argument source quality), TAQ (Total Argument quality), and FAA (Final Argument Assessment), under consideration of the independent variables G (Gender), E (Ethnicity), PE (Pell Eligibility), and FG (First Generation), the following findings emerged:

Argument Content Quality among demographic groups. The ACQ mean was 2.19 out of the maximum score 5, based on the six-level grading scale of very low (approaching), low (proficient), medium (emerging), high (advanced), very high (advanced), and exceptional (advanced). The mean score falls on level 2 (medium). Female scored higher than male students (2.29 vs 2.00), Latinx scored higher than African American students (2.18 vs 2.00), First Generation students scored slightly lower than non-First Generation students (2.21 vs 2.24), and Pell Eligible scored lower than non-Pell Eligible students (2.11 vs 2.36).

Argument Sourcing Quality among demographic groups. The ASQ mean value was 1.06 out of a maximum score of 5, which indicates that students in the sample fall on the low level. Male scored higher than female students (1.12 vs 1.03), African American scored higher than Latinx students (1.30 vs 1.03), First Generation students scored higher than non-First Generation students (1.15 vs.92), and Pell Eligible students scored higher than non-Pell Eligible students (1.11 vs.93).

Total Argument Quality among demographic groups. The TAQ mean value for the sample was 3.26, out of a maximum of 5, which indicates that students in the sample fall on the high level. Female scored higher than male students (3.32 vs 3.12), African American scored higher than Latinx students (3.30 vs 3.22), First Generation students scored higher than non-First Generation students (3.36 vs 3.16), and Pell Eligible students scored lower than non-Pell Eligible students (3.22) vs Pell Eligible (3.29).

Final Argument Assessment among demographic groups. The FAA mean value for the sample was .43, out of a maximum of 1; female scored higher than male students (.49 vs .33), Latinx scored higher than African American students (.44 vs .35), First Generation students scored lower than non-First Generation students (.40 vs .47), and Pell Eligible students scored lower than non-Pell Eligible students (.41 vs .46).

Descriptive statistics of the highest-scoring students. Among the twenty-two students who scored the highest (11-14) on TU were females and First Generation students, African American students (except for three who were Latinx), and Pell Eligible and non-Pell Eligible students at equal halves. At the highest level (14), students were interested and somewhat familiar with the topic; they scored medium on ACQ, answered FAA correctly, and scored at the advanced level on RC and CK. At level 13, the students were First Generation, of mixed gender, and also interested in the topic; they scored correctly on FAA and had a medium ACQ score, while their RC was emerging to proficient, and their CK was advanced. At level 12, female students dominated, mostly First Generation; they were interested in the topic, but not familiar with it. They scored medium to high on ACQ, answered FAA correctly, and had advanced CK. At level 11 were mostly female students, somewhat interested to interested, and somewhat familiar to not familiar; their ACQ score ranged from medium to high, and their CK was advanced (Table 17).

Table 17

Students who scored highest on TU (11-14)

Number of Students	TU score	TU Level	Demographic Characteristics	Other Emerging Qualities
2	14	Advanced	Female, FG	Very interested to interested, somewhat familiar, ACQ medium, FAA correct, RC advanced, CK advanced
3	13	Advanced	Male/Female, FG,	Very interested to interested, ACQ medium, FAA correct, CK advanced, RC emerging to proficient
8	12	Advanced	Female, FG	Very interested to interested, not familiar, ACQ medium to high, FAA correct, CK advanced
9	11	Advanced	Female	Somewhat to interested, somewhat to not familiar, ACQ medium to high, CK advanced

Notes. See Appendices for the full TU max and TU min tables with individual students' characteristics.

Descriptive statistics of the lowest-scoring students. The seventeen students who scored the lowest on TU were of mixed gender, and most of them were African American, except for one who was Latinx. They were mostly First Generation students, with a digital reading time of less than five years. At the lowest level recorded (4), students incorrectly responded to FAA, but their RC score was advanced, and their CK score was proficient; they were non-Pell Eligible students (Table 18).

Table 18

Students who scored lowest on TU (4-6)

Number of students	TU score	TU Level	Demographic Characteristics	Other Emerging Qualities
2	4	Proficient	Female, FG, PE (0) not eligible, African Americans	DRT 5-10 yrs., ACQ medium, FAA incorrect, RC advanced, CK proficient
4	5	Proficient	50/50 Male/Female, Mostly PE (0) not eligible, FG, African Americans	DRT mostly >10 yrs., ACQ medium, CK proficient, ASQ low, TAQ high
11	6	Proficient	Mostly Female 8/11, FG, African Americans	Mostly interesting, somewhat familiar, RC proficient, CK emerging, ACQ medium, mostly DRT <5 yrs.

Notes. TU scale (approaching 0-1, proficient 2-6, emerging 7-9, advanced 10-16), RC scale (approaching 0, proficient 1, emerging 2, advanced 3), CK scale (approaching 0-1, proficient 2-4, emerging 5-6, advanced 7-12), ASQ scale (very low 0, low 1, medium 2, high 3, very high 4, exceptional 5), ACQ scale (very low 0, low 1, medium 2, high 3, very high 4, exceptional/advanced 5), TAQ scale (very low 0, low 1, medium 2, high 3, very high 4, exceptional/advanced 5), FAA scale (0,1).

At level 5, most students were non-Pell Eligible, First Generation, and their digital reading time was less than 10 years; their ACQ was medium, ASQ low, thus TAQ high, and most of them were African Americans. Finally, at level 6, most students were female (8/11), mostly interested in and somewhat familiar with the topic. They scored medium on ACQ, and high to very high on TAQ, while their RC was proficient, their CK emerging, and their digital reading time less than five years. Most of them were African American (Table 18).

This section has discussed the descriptive statistics of the dependent variables and background variables, while the following section on inferential statistics will highlight the statistically significant relationships between these variables.

Inferential Statistics

The inferential statistics added valuable information, showing which values are statistically significant. Similarly, as with the descriptive statistics, we ran inferential statistics to find out how independent variables (Gender, Ethnicity, First Generation, Pell Eligibility) affected dependent variables (TU, RC, CK, ACQ, ASQ, TAQ, and FAA). The following paragraphs and tables present the findings from the inferential statistics analyses.

Independent t-test results on total understanding and argument quality in gender groups. The inferential statistics revealed that, when it comes to gender, none of the dependent variables, except for ACQ (Argument Content Quality) and FAA (Final Argument Assessment), emerged as statistically significant (Table 19, and Table 20).

Table 19

Inferential Statistics for Total Understanding, Whole Sample by Gender

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	t	df	Sig 2-sided
Reading Comprehension	4.306	.040	-.904	121.767	.368
Content Knowledge	.001	.979	-.098	154	.922
Total Understanding	.422	.517	-.258	154	.797

According to the Independent t-Test of Equality of Means, there was no significant difference in performance between male and female students in terms of Reading Comprehension, Content Knowledge, and Total Understanding in the whole sample, nor when it came to paper and digital condition. However, the ACQ and FAA scores are significant, which means that female students performed better than male students on Argument Content Quality and Final Argument Assessment (Table 20).

Table 20

Inferential Statistics for Argument Quality and Assessment, Whole Sample by Gender.

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	T	df	Sig 2-sided
Argument Content Quality	3.374	.068	2.147	153	.033
Argument Source Quality	.355	.552	-.489	153	.625
Total Argument Quality	.125	.724	.928	153	.355
Final Argument Assessment	13.521	<.001	1.982	108.284	.050

Note. ACQ and FAA were at significant levels.

Next, we looked at gender groups and how they scored on topic interest, topic familiarity, and digital reading time (Table 21).

Table 21

Inferential Statistics for Topic Interest, Familiarity and Digital Reading Time, Whole Sample by Gender

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	T	df	Sig 2-sided
Topic Interest	3.262	.073	-.955	152	.341
Topic Familiarity	1.155	.284	-.852	151	.396
Digital Reading Time	.186	.667	-1.862	151	.064

Note. Digital reading time was nearing significance.

Independent t-test results on total understanding and argument quality in ethnicity groups. The inferential statistics also included two ethnic groups, African American students and Latinx students (Table 22).

Table 22

Inferential Statistics for Total Understanding, Whole Sample by Ethnicity

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	t	df	Sig 2-sided
Reading Comprehension	.105	.746	.509	139	.612
Content Knowledge	.028	.868	.575	139	.566
Total Understanding	.198	.657	.707	139	.481

Similar findings emerged for the different ethnic groups in terms of argument quality assessment with dependent variables ASQ, ACQ, TAQ, and FAA (Table 23).

Table 23

Inferential Statistics for Argument Quality and Assessment, Whole Sample by Ethnicity

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	t	df	Sig 2-sided
Argument Content Quality	.572	.451	-.920	138	.359
Argument Source Quality	.241	.624	1.096	138	.275
Total Argument Quality	.052	.820	.262	138	.793
Final Argument Assessment	4.249	.041	-.773	26.021	.446

Finally, students' topic interest, familiarity, and digital reading time also were insignificant among the two ethnicities (Table 24).

Table 24

Inferential Statistics for Topic Interest, Familiarity, and Digital Reading Time, Whole Sample by Ethnicity

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	t	df	Sig 2-sided
Topic Interest	4.289	.040	-.695	138	.488
Topic Familiarity	.567	.453	.592	24.218	.559
Digital Reading Time	3.065	.082	1.462	26.472	.155

Independent t-test results on total understanding and argument quality in First Generation and non-First Generation students. The inferential statistics of the whole sample revealed statistically significant values for RC, but not for the variables TU or CK, among the First Generation students (Table 25).

Table 25

Inferential Statistics for Total Understanding, Whole Sample by First Generation versus Parents some College, and Graduated from College

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	t	df	Sig 2-sided
Reading Comprehension	.971	.326	-1.965	141	.051
Content Knowledge	1.084	.300	.438	141	.662
Total Understanding	.038	.846	-.437	141	.663

Note. RC reached significant level.

There was a significant difference in performance between First Generation students and students with a Parent with some college education or college graduate when it came to the variable of Reading Comprehension (RC), such that the latter performed better than First Generation students.

Next, we also ran the Independent t-test on the variables related to argument quality, for ACQ, ASQ, TAQ, and FAA (Table 26).

Table 26

Inferential Statistics for Argument Quality and Assessment, Whole Sample by First Generation vs Parents some College and Graduated from College.

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	t	df	Sig 2-sided
Argument Content Quality	.024	.878	-.189	140	.851
Argument Source Quality	.321	.572	1.252	140	.213
Total Argument Quality	3.836	.052	.966	127.437	.336

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	t	df	Sig 2-sided
Final Argument Assessment	1.856	.175	-.864	140	.389

Finally, topic interest, topic familiarity, and digital reading time were analyzed for statistical significance (Table 27).

Table 27

Inferential Statistics for Topic Interest, Familiarity, and Digital Reading Time, Whole Sample by First Generation vs Parents some College and Graduated from College.

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	t	df	Sig 2-sided
Topic Interest	.599	.440	.897	140	.371
Topic familiarity	.808	.370	.388	139	.699
Digital Reading Time	5.307	.023	-.219	120.160	.827

Independent t-test results on total understanding and argument quality among Pell Eligible and non-Pell Eligible students. The inferential statistics for the whole sample revealed that RC was almost statistically significant (Table 28), such that non-Pell Eligible students scored significantly higher than Pell Eligible students.

Table 28

Inferential Statistics for Total Understanding, Whole Sample by Pell Eligible

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	t	df	Sig 2-sided
Reading Comprehension	6.335	.013	1.913	121.316	.058
Content Knowledge	2.443	.120	-.818	155	.415
Total Understanding	.847	.359	.080	155	.937

Note. RC was approaching statistical significance.

Next, the Independent t-test scanned Pell Eligible students on argument quality (Table 29, and Table 30).

Table 29

Inferential Statistics for Argument Quality and Assessment, Whole Sample by Pell Eligible

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	t	df	Sig 2-sided
Argument Content Quality	.009	.923	-1.874	147	.063
Argument Source Quality	1.699	.194	1.088	147	.278
Total Argument Quality	3.611	.059	-.343	147	.732
Final Argument Assessment	1.082	.300	-.590	146	.556

Notes: None of the variables are significant in the whole sample.

While Argument Quality assessment revealed no significant findings, the Argument Content Quality approached significance level at .63.

Table 30

Inferential Statistics for Topic Interest, Familiarity, and Digital Reading Time, Whole Sample by Pell Eligible

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	t	df	Sig 2-sided
Topic Interest	.327	.569	.202	153	.840
Topic familiarity	.212	.646	-1.237	152	.218
Digital Reading Time	3.458	.065	1.048	152	.296

Notes: None of the variables are significant in the whole sample.

It should be noted that we conducted a number of analyses that yielded non-significant findings. We compared the responses of demographic groups, e.g., male vs female, under the paper only condition, and again under the digital only condition. In addition, we estimated regression equations predicting the key outcomes with the background independent variables, including interest and prior knowledge. In all these analyses, few statistics yielded significant results. The exceptions, some of which might represent Type I errors, are reported next.

Independent t-test statistically significant results and respective descriptives for students under paper and digital conditions. The inferential statistics for students tested under two conditions revealed statistically significant results for FAA, ASQ, RC, and TF. They are presented in the following, by Gender, by Ethnicity, by First Generation and Pell Eligible student groups.

Table 31

Descriptive statistics for Topic Interest, under Digital Condition by Gender

Variable	Gender	N	Mean	Std. Deviation
Topic Interest	Female	55	1.38	.871
	Male	24	1.71	.908
Topic Familiarity	Female	55	.53	.813
	Male	23	.78	1.043
Digital Reding Time	Female	55	.71	.712
	Male	23	1.04	.767

Notes. In digital condition none of the variables were significant, except DRT that was approaching significance, in which case male scored higher.

Under digital condition, there was no difference between gender when it came to students reporting topic familiarity and topic interest. However, the digital reading time was higher among male students at the level approaching significance.

Table 32

Inferential Statistics for Topic Interest, under Digital Condition by Gender

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	t	df	Sig 2-sided
Topic Interest	.037	.848	-1.513	77	.134
Topic Familiarity	1.338	.251	-1.161	76	.249

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	t	df	Sig 2-sided
Digital Reading Time	.144	.705	-1.849	76	.068

In digital condition none of the variables were significant, except DRT, which was approaching significance at .068.

Table 33

Inferential Statistics for Argument Quality and Assessment, Paper Condition by Ethnicity

Variable	Ethnicity	N	Mean	Std. Deviation
Argument Content Quality	African American	15	1.87	.990
	Latinx	53	2.11	.776
Argument Sourcing Quality	African American	15	1.47	.834
	Latinx	53	.91	.904
Total Argument Quality	African American	15	3.33	1.447
	Latinx	53	3.02	1.293
Final Argument Assessment	African American	15	.27	.458

Under the paper condition, African American students scored significantly higher on argument sourcing quality (ASQ) than Latinx students.

Table 34

Inferential Statistics for Argument Quality and Assessment, Paper Condition by Ethnicity

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	t	df	Sig 2-sided
Argument Content Quality	.732	.395	-1.020	66	.311
Argument Source Quality	.000	.995	2.156	66	.035
Total Argument Quality	.295	.608	.810	66	.421
Final Argument Assessment	2.264	.137	-.656	66	.514

Note. Argument Sourcing Quality was at significant level.

Under the paper condition, African American students' higher performance was statistically significant (.035) for ASQ. Furthermore, under the paper condition non-First Generation students scored higher (1.54) on reading comprehension than First-Generation students (1.13).

Table 35

Descriptive Statistics for Total Understanding, Paper Condition by First Generation

Variable	First Gen/Parent College	N	Mean	Std. Deviation
Reading Comprehension	First Generation	45	1.13	.869
	Parent College/Grad	24	1.54	.779
Content Knowledge	First Generation	45	7.64	1.747
	Parent College/Grad	24	7.33	1.685
Total Understanding	First Generation	45	8.78	2.033
	Parent College/Grad	24	8.88	2.028

Notes. Reading Comprehension is the sum of Q1-Q4_RC, Content Knowledge is the sum of Q5-Q16, while Total Understanding (TU) is the sum of RC and CK.

Under the paper condition, the difference in performance between First Generation students and students whose parents had graduated from college were not significantly different, except for RC, which was approaching significance (.058).

Table 36

Inferential Statistics for Total Understanding, Paper Condition by First Generation

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	t	df	Sig 2-sided
Reading Comprehension	.130	.720	-1.926	67	.058
Content Knowledge	.255	.615	.713	67	.478
Total Understanding	.002	.967	-.189	67	.850

Under the paper condition, the scores among Pell Eligible and non-Pell Eligible students were not significantly different, except for topic familiarity, which was statistically significant (.041).

Table 37

Descriptive Statistics for TI, under Paper Condition by Pell Eligible

Variable	Pell Eligible	N	Mean	Std. Deviation
Topic Interest	Pell Eligible	42	1.19	.833
	Non-Pell Eligible	33	1.33	.736
Topic Familiarity	Pell Eligible	42	.76	.692
	Non-Pell Eligible	33	.42	.708
Digital Reading Time	Pell Eligible	42	.69	.680
	Non-Pell Eligible	33	.79	.696

Under the paper condition, Pell Eligible students scored higher (.76) than non-Pell Eligible (.42) students in terms of topic familiarity.

Table 38

Inferential Statistics for TI, under Paper Condition by Pell Eligible

Variable	Levene's Test for Equity of Variances		Independent t-Test for Equity of Means Significance		
	F	Sig	t	df	Sig 2-sided
Topic Interest	.419	.520	-.775	73	.441
Topic Familiarity	.027	.870	2.076	73	.041
Digital Reading Time	.053	.818	-.609	73	.544

Pell Eligible students tested under the paper condition were more familiar with the topic than students whose parents had attended some college or graduated from college. The following section will summarize the quantitative results, followed by the qualitative findings.

Summary of Quantitative Results

According to the Independent T-test for the equality of means on the whole sample for Argument Content Quality (ACQ), female students performed better than male students. Similarly, on Final Argument Assessment (FAA), female students did better than male students. Students whose parents had attended college or graduated from college performed better on Reading Comprehension (RC) than First Generation students. When testing for Reading Comprehension among non-Pell eligible and Pell eligible student groups, the first group scored higher than the latter.

For Total Understanding (TU), the highest mean score among all four demographic groups was recorded specifically for African Americans, while the lowest score was seen with Latinx students. On RC, the highest mean score was obtained by the non-First Generation student group, and the lowest by First Generation students. On CK, the highest score was reported specifically by African Americans, and the lowest by the non-Pell Eligible group. The TU variable at the advanced level of 11-14 was achieved by twenty-two students, which accounts for 14% of the sample, while the lowest level, proficient in the point range of 4-6, was scored by seventeen students (11%) out of the 157 students in the sample.

In terms of the argument quality variables, the highest score for ACQ was reached by female students, and the lowest by African Americans. On the ASQ variable, African American students scored the highest while non-First Generation students scored the lowest; also, female and Latinx students achieved relatively low scores. On the TAQ variable, the highest number of points was scored by First Generation students, while the lowest number was scored by male students. Finally, on FAA, the highest mean score was obtained by female students, scoring highest among all groups, while male students scored the lowest, and so did African Americans.

Based on the descriptive and inferential statistics of the sample N=157 of undergraduate students, for the dependent variables TU (total understanding), RC (reading comprehension), CK (content knowledge), under consideration of the independent variables G (Gender), E (Ethnicity), FG (First Generation), and PE (Pell Eligibility), the following findings emerged:

The TU mean was 8.92 out of a maximum score of 16, based on the four-level grading scale (approaching, proficient, emerging, advanced). This falls on level 3 (emerging); female scored higher than male students (8.95 vs 8.97), African American scored higher than Latinx students (9.19 vs 8.86), First Generation students scored lower than non-First Generation students (8.97 vs 9.12), and Pell Eligible students scored slightly lower than non-Pell Eligible students (8.91 vs 8.94).

The RC mean value was 1.24 out of the maximum score of 3, which indicates that students in the sample fall on the proficient level (low level), female scored higher than male students (1.27 vs 1.15), African American scored higher than Latinx students (1.29 vs 1.15), First Generation students scored lower than non-First Generation students (1.13 vs 1.41), and Pell Eligible students scored lower than non-Pell Eligible students (1.13 vs 1.38).

The CK mean value for the sample was 7.69 out of a maximum of 12, which indicates that students in the sample fall on the advanced level; female scored higher than male students (7.68 vs 7.71), African American scored higher than Latinx student (7.90 vs 7.67), First Generation students scored higher than non-First Generation students (7.84 vs 7.71), and Pell Eligible students scored higher than non-Pell Eligible students (7.78 vs 7.55).

Based on the descriptive and inferential statistics of the sample N=157 of undergraduate students, for the dependent variables ACQ (argument content quality), ASQ (argument source quality), TAQ (Total Argument quality), and FAA (Final Argument Assessment), under consideration of the independent variables G (Gender), E (Ethnicity), PE (Pell Eligibility), and FG (First Generation), the following findings emerged:

The ACQ mean was 2.19 out of a maximum score of 5, based on a six-level grading scale—ranging from very low (approaching), low (proficient), medium (emerging), high (advanced), very high (advanced), to exceptional (advanced). Hence, the mean falls on level 2 (medium). Female students scored higher than male students (2.29 vs 2.00), Latinx scored higher than African American students (2.18 vs 2.00), First Generation students scored lower than non-First Generation (2.21 vs 2.24), and Pell Eligible students scored lower than non-Pell Eligible students (2.11 vs 2.36).

The ASQ mean value was 1.06, out of the maximum score of 5, which indicates that students in the sample fall on the low level. Male scored higher than female students (1.12 vs 1.03), African American scored higher than Latinx students (1.30 vs 1.03), First Generation students scored higher than non-First Generation students (1.15 vs.92), and Pell Eligible students scored higher than non-Pell Eligible students (1.11 vs .93).

The TAQ mean value for the sample was 3.26, out of a maximum of 5, which indicates that students in the sample fall on the high level; female scored higher than male students (3.32 vs 3.12), African American scored higher than Latinx students (3.30 vs 3.22), First Generation students scored higher than non-First Generation students (3.36 vs 3.16), and Pell Eligible students scored higher than non-Pell Eligible students (3.22 vs 3.29).

The FAA mean value for the sample was .43, out of a maximum of 1; female scored higher than male students (.49 vs .33), Latinx scored higher than African American students (.44 vs .35), First Generation students scored lower than non-First Generation students (.40 vs.47), Pell Eligible students scored lower than non-Pell Eligible students (.41 vs.46). Under digital condition, male students reported higher Digital Reading Time (1.04 vs .71) at the level approaching significance. In general, higher scores were obtained by female First-Generation African American students, who also did well on RC and TAQ.

The following section will present the qualitative results from a series of interviews with the students.

Qualitative Results

Results of the Semi-Structured Interviews

This overview presents the results from the qualitative data collection of students' perception at an institution of higher education in Southern California, with a predominant enrollment of people of color. The qualitative part of this study involved semi-structured interviews comprised of 11 open-ended questions focusing on students' preparedness for college and their experience with learning from electronic or paper texts. The aim of this qualitative part of the study was to gather descriptive data so as to better understand students' perception about their college preparedness and the aspects that drive their motivation towards graduation. This narrative approach provided an opportunity to collect students' educational experiences grouped by emerging themes, which were compared and then explained based on relatedness, competence, and autonomy—that is, the components of Self-Determination Theory.

Students were personally approached, in or outside class, and invited by the researcher and their professors to participate in the interview. The interviews took place either in person or online; either way, they were recorded with Zoom. Prior to the conversation, the students signed a consent form; however, their names are kept confidential. The aim of this qualitative section is to complement the quantitative survey data collected earlier. The quantitative section relied on a questionnaire based on reading and science literacy frameworks and provided variables for statistical analyses of reading comprehension of a difficult scientific text presented on paper or screen. In contrast, the qualitative section relied on the students' responses to interview questions. Recorded narratives were transcribed by using transcribe.app. Then, the most relevant and complete elaborations on the topics studied were selected and grouped into emerging themes that reflect the changes mandated by the recently implemented Assembly Bill (AB) 1705 Bill, while the Self-Determination Theory guided the explanation of these emerging themes—that is, relatedness, competence, and autonomy (Deci, and Ryan, 1985, 2013). The interview was comprised of the questions listed in Appendix A.

Interviews and emerging themes. During the interviews, students responded to questions about their preparedness for college-level classes and their experience reading from paper and digital screen. Thus, the first theme that emerged from student's narratives is competence.

Feeling competent. Felling competent due to adequate high school preparation. If the ability to be effective or competent is not met, students feel inadequate (Earl et al., 2017). Competence contributes to students' well-being and intrinsic motivation.

During the interviews, students expressed satisfaction about their preparedness for college-level courses due to their high school's strong emphasis on college access, as well as offerings of Advance Placement (AP) college courses. Below are selected quotes from students' narrative responses:

> I do feel prepared. That’s mainly because I did take a lot of AP classes in HS, so just by that metric I kind of got good idea and a lot of experience about what (…) college-level class would actually be.

> Um, I do feel prepared for college-level courses. I did take an AP class in high school my senior year. It was only one AP class, but it was composition literature. (…) But I felt ready for college-level classes (…)

Feeling competent due to low reading expectations. Most of the participants felt at ease and not stressed by the number of classes that they were required to take during one semester, and they did not seem to be specifically overwhelmed by the amount of reading and writing. Students said:

> I haven’t had too much reading to do, and when I do get readings, it’s not like something that is (…) pressured or anything.

> (...) I didn’t experience that much reading because it was more research and you just (...) you can rent the books, but there wasn’t that much reading in any of my classes that I took while I was here. It was more research, taking notes, try to answer research questions to the best of your ability, and use these textbooks as a resource.

Feeling competent due the instructor-provided lecture notes and slides. Students often mentioned that there was no excessive reading because most classes relied on lecture notes and power point slides.

> We mostly get Power Points. Most of the concepts are on Power Points these days. It's only a few classes that they still expect us to read the textbook.

> I think for my classes it is a good amount (of reading). Most of my classes are in person lecture. So, there is not a lot of reading that I need to do, as far as reading from a textbook or articles, but there is enough reading for me to supplement the actual class that we do in person, so, yes, I think it is about right.

Feeling competent due to strong educational foundations and writing skills. A recurring theme in the students' narratives is their emphasis on writing, math, and college preparation in high school; moreover, they mentioned offerings of AP courses for prospective college students as well as counseling services. One student went to the extent to compare his experiences in the public high school to the newly opened private charter school to which he transferred. He explained how opening the small charter school triggered an improvement of the adjacent K-12 public schools, which now had to compete for students. He stated that the charter school focused on the hands-on application of math and science to solve real-life problems, such as managing a budget, or participating in elections.

> (…) at our school (the charter school), we had a growth mindset, you know, we played with robots; we had an economics class where we actually learned, you know, this is how

you do taxes. In economics class at a public high school, they don't teach you how to do taxes. But we actually went through a simulated, you know, tax refund in our economics class, and we made a budget, and we talked about it in our civics class. We didn't just learn, oh, this is how the government works. (…) We learned this is how politicians get funding and this is how you vote, and this is how you decide, you know, how you decide what you want to vote for. We learned a lot more life, the real-world stuff.

So, yes, I felt very prepared for college-level writing before I got to college. And since then, I've taken multiple English classes and composition classes. I have not gotten to the senior-level English, to the senior-level composition courses, which I'm expected to take next semester. I'm expecting they're going to be very challenging, but I hope I'm well prepared for that. (…) Well, I'm a big reader. It's just right from me.

Most students felt prepared for college English and math, for these were core subjects covered in high school; moreover, most did not feel surprised by the level of requirements in their courses, because they had taken advantage of the AP college courses offered in their high school.

I do feel prepared. (...) one of my strongest (...) characteristics, (...) when it comes to (...) college classes, a college course requires me to write an essay, I would be a hundred percent prepared to write something, especially since I took that AP course in high school, that definitely helped out with, you know, trying to write essays in the timeframe (...)

Not feeling prepared for college-level science. Most of the students appreciated their high school focusing on college preparation, in particular writing and math; thus, students seemed to be competent and confident about successfully completing the degree. Although many expressed an interest in science, there were several students who indicated that college science classes were much more advanced than what they were prepared for:

> I think I felt least prepared for a lot of science courses. Just because they have never been something I wanted to dig into or (was) motivated to try to learn and engage with. The science courses I am taking this semester have been hands-on, pretty intense, as far as the level from which I am supposed to learn the material (...) with science and Earth Science, the stuff that you learn is so much more qualitative than, like, math. Like, math equations have one answer, numerical answer, whereas in the Earth Science field the answer to something, it's very complex and descriptive, and it's not one-dimensional. It has so many dimensions to it… certain aspects in Earth Sciences and Physical Sciences, I have a hard time wrapping my head around.

Feeling competent due to good writing skills. Overall, students were content with how well their high school had prepared them for college-level writing and math. Most students were confident about their writing and math skills, but only a few felt confident about their preparedness for college-level science. When it came to reading, students generally did not find the amount of college reading too excessive; quite to the contrary, one student said that there was not enough reading assigned.

Several Latinx students indicated pride in having good English writing skills. Apparently, English preparation has been improving for Latinx students; however, physical sciences and math might not yet have caught up with this improvement. It will take time before science and math preparation will match preparation in language skills. Among the interviewees, there were only few students who felt comfortable with their skills in the physical sciences. Out of 16 students, only three chose a science major—that is, physics or astrophysics, microbiology, and geography—but only two felt comfortable with math.

Relatedness. Students' ability to relate to a topic, environment, or culture fulfills their psychological need for connection (Earl et al., 2017). Familiarity with content or environment gives them comfort and provides an external motivation to pursue their goals. Students in general felt a sense of belonging to the university and its culture. They did not feel stressed by the number of the required courses, nor by the amount of reading and writing assigned in their classes. The university atmosphere and culture suited the interviewees, for there was no student who felt an outsider or not welcomed. Those who expressed their feelings about campus life described a sense of belonging, enjoyed the friendly atmosphere, and felt proud to be a part of this community. The interviewees were students who came either from neighboring communities, or from other states and regions, such as New York, the San Francisco Bay Area, Orange County, or Los Angeles. A few students—and particularly those who had not done well in high school—expressed appreciation that they had been accepted. During the interview, one student honestly admitted that this was the only university that had accepted him:

> So, before I chose a university, I made sure what was the diversity? How many, like, what's the percentage of, like, people of color? So, my high school was linked up to UCSD (...). So, once I was checking UCSD, I saw it didn't have that much, like, diversity of people of color. So, once I looked at this one, they had liked a lot of, uh, people of color, and the majority of the students here are Hispanic.

> And to that extent, I do feel that, yes, I do feel like I belong here. I do feel, like, there are opportunities presented towards me, and that people do their job to the highest extent and go out of their way to make sure that we know that we are supported.

> Uh, the transition was okay. It was about what I expected, but actually I feel like people are a lot nicer and more open and welcoming than I was expecting. So, I'm not sure if (this is about a) sense of belonging per se, but I'm really glad of people who I've met so far. Okay (...) I am a part of a club for my major and the honor society for my major as well, but no study groups.

Autonomy. Feeling independent corresponds to students' internal motivation, and if the need for freedom were not met, students would feel pressured to behave in a specific way. Apparently, most interviewed students were satisfied with the opportunity to choose or switch majors in order to suit their interests and strengths. Students recognized the availability of educational funds, housing assistance, and other services alleviating shortcomings and specifically directed to students from underrepresented groups. Students clearly expressed their joy about the resources, services, opportunities, and freedom granted by the university. After transferring from high school, students recognized and appreciated the weight of responsibility for their own educational journey, but they also mentioned that they could control their schedule, select classes, and choose as many units per semester as they desired:

> (...) personally, I have never felt like the school is what defines me. I think what defines me is what I seek to do personally, so I don’t necessarily feel like I am a part of the school community. I just go to school to do what I want to do. If that makes sense.

> Well, the transition (...) I felt like it was a pretty easy transition. But, I guess, for the social part, I did struggle because I transferred from one university to another within my first two years of college (...) so, I haven’t made too many friends, but only because I’m focusing on work, and I haven’t had social life at the university.

Other noteworthy themes relevant to this study consisted of experiences with STEM and with reading from paper and digital texts, the impact of AB 1705, and prevalent study majors.

Emerging Students’ Majors. Most of the students interviewed chose those majors about which they felt passionate and in which they felt most competent. Only one student was not sure what major to pursue and, therefore, selected the first major on the list. Two students decided to switch their major during their program. The following are examples of students’ responses about the journey towards their desired major:

Biology. Yes. And then now that I'm here, you know, there's a good balance between physical geography, and I've taken some computer, computer science classes with Dr. Ch (…) of course, the World Ocean Biology class. I have a good foundation on those social sciences. I'm not going in depth with those, but I have good basics.

Human geography. I've taken one of those courses. So, I feel I have a good foundation of both physical and social science.

(...) I was thinking about, like, arts stuff, but I didn't know what, and one day I was just, like, lying in bed listening to music, and I was, like, I like sound and music. So, I went into audio engineering, and that's how I chose this major.

Experience with Science. The theme of science experience is considered separately, since it is of interest in terms of how many interviewees felt competent and motivated to pursue this field of study. The statements below present selected findings:

The classes I feel most prepared for are math and writing, like literature, just because that was the bulk of the classes that I took in high school. The bulk was math, literature, and classes pertaining that those areas of study. Whereas science was never something I really took on very much.

(...) Physics is a bit of a challenge… I would say physical science, I have a good foundation, and then fortunately, I'm lucky it's the classes I'm really interested in. For

> example, I'm taking World Ocean with Dr. S. (...) So fortunately for me, you know, a lot of the stuff in the textbook is very complex. So fortunately for me, he has very detailed Power Points, and his lectures are recorded. It depends on what type of learner you are, I'm a sort of half auditory, half visual learner.

> I feel most prepared for those STEM classes, such as this Earth 101 class that I'm currently taking now. I wouldn't feel prepared for it, but I think that's only because of like how my work ethic is. I work better in certain areas, and with the math and sciences, that would be something that I would need extra help to, you know, do well in that class.

In general, students expressed confidence in their preparedness for college, in particular concerning English; however, there were fewer who felt prepared in math or science. Students expressed their appreciation about their high school's preparation and AP course offerings. They also highlighted the welcoming university environment, the support services, and resources available, and finally having the opportunity to socialize.

Qualitative Study Summary. This qualitative study provided valuable insights into students' preparedness for college-level writing, reading, math, and sciences.

The interview responses assisted the researcher in analyzing the narratives through the lens of the components of basic human psychological needs as per the Self-Determination Theory (Daci and Ryan, 1985, 2013), which are competency, relatedness, and autonomy. Thus, students' responses indicating their feeling competent due to their HS preparation for college level math, English and science was categorized in the first group: competency, next students indicating the feel of belonging because of the students' population and resources for students of color and underrepresented students was categorized as relatedness, and finally autonomy category included students' responses about their agency navigating through college towards graduation.

Students talked about counseling services at the university which help them navigate towards graduation, in spite of their challenges with math, for example. One student talked about how he switched between universities, simply because he couldn't successfully pass math and advance towards an engineering degree. However, once he arrived at this university, he was advised to choose a geography major where he was more likely to succeed. He also elaborated on the resources specific to the needs of underrepresented groups (such as Latinx, African American, and LGBTQ) and praised the access to supportive resources, educational planning services, and academic intervention.

Only one student talked about a poor public high school experience and about the transfer to the private charter school which he, as a matter of fact, genuinely enjoyed attending. He explained having access to the resources that the school had to offer, thanks to the financial support provided by the adjacent space industry. He also explained that the opening of one small charter school triggered the improvement of public schools in the same neighborhood.

Both the quantitative and the qualitative parts of this study share one common result: over time, the gap in students' performance diminishes at the age of 17 and older, since, in spite of students' differences in earlier years, they perform about the same. The quantitative part of this study is grounded in international standardized assessments (such as PISA and TIMSS). PISA and TIMSS findings also concluded that the performance gap between students, which is prevalent at a younger age in K-12, diminishes quickly as students become 17 and older.

The qualitative part of this study is grounded in the Self -Determination Theory and was also supported by the new legislation derived from the results of the standardized assessments, AB 1705, which mandates universities to place students in transfer courses not in remedial courses, for if even when they lack skills, the studies show, they would eventually catch up; moreover, students do not need to take courses that are not required for their major.

In effect students did not feel stressed from too much study, or too much reading or writing. They expressed enjoyment with the university recourses, and the amount of assigned work.

Also, in the students' responses, a recurring theme was not being overwhelmed by the number of required classes, because of the reduced number of units needed for a degree. This has also decreased the time towards completion or transfer, and it will reduce the cost of college.

CHAPTER V Discussion and Conclusion

Research Question

What is the level of undergraduates' preparedness for college when assessed on the basis of their comprehension of scientific texts under the conditions of students' background (demographics) and students' perception (topic familiarity and knowledge), with consideration of paper versus digital texts?

Key Findings

The predominant population included in this study are students of color, specifically Latinx and African Americans. The majority of students in the sample were First Generation students; about one-quarter had parents with some college, and even fewer had parents who graduated from college. In the sample, about half of the students were Pell Eligible.

The descriptive statistics of the entire sample concerning RC, KC, and TU rated students' RC as low, while CK and TU were low and medium, at about 50 percent, and only 2 percent of students scored at the high level. These findings align with the 2017 OECD and PISA Brief, stating that "in the United States only 12% (1/6) of adults scored at the highest level on standardized tests on the basic literacy skills, which means they have the weak literacy skills" (PISA, 2017; OECD).

Among the gender groups, the higher mean value was scored by females, on both the TU and RC assessments, but for CK the higher mean value was scored by males. Female students scored higher than male students on ACQ and TAQ, except for ASQ; as for the latter, male students scored higher than female students. Also, male students scored higher on TI and TF.

In terms of the argument quality assessment—such as ACQ, TAQ, and ASQ—in the entire sample students performed better on TAQ than ACQ and ASQ. These results coincide with prior findings by Salmeron (2018), stating that "students at all educational levels often disregard source information and pay attention only to document content" (Bråten et al., 2018; Salmerón, Gila, and Bråten, 2018, p. 25). However, in contrast to the findings presented by Bråten, Strømsø, and Salmerón (2011), prior knowledge (TF) did not have a direct effect on ACQ, but rather a reverse relationship (starting with being very familiar with the topic, students' mean score was 1.67, familiar 2.23, somewhat familiar 2.08, not familiar 2.28). Topic familiarity had a positive direct effect on students' ASQ, and the effect was statistically significant: those who were familiar scored 1.15, which was less than the 2.33 scored by the group who was very familiar. From very familiar to not familiar, mean scores on ASQ were: 2.33, 1.15, 1.06, and .98, respectively. Similarly, for TAQ there was a positive direct relationship, from very familiar scoring 4.00, familiar 3.38, somewhat familiar 3.14, to unfamiliar 3.26, respectively. FAA only partially coincided with topic familiarity, such that very familiar students scored the highest, but familiar student did not score higher than somewhat familiar or unfamiliar students.

Topic interest did not have much of an effect on ACQ, ASQ, TAQ, or FAA. From the topic being very interesting to not interesting at all, the mean scores were rather reversely related, as follows: ACQ [reverse trend]: 2.13, 2.18, 2.25, 2.05; ASQ [no specific trend; it is significant that somewhat interesting at .88 scored lower than interesting at 1.42]: 1.13, 1.42, 0.88, 0.82; TAQ [no specific trend; it is significant that somewhat interesting at 3.13 scored lower than interesting 3.60]: 3.27, 3.60, 3.13, and 2.86; and FAA [reverse]: 0.40, 0.42, 0.49, and

0.32. McCrudden, Stenseth, Bråten, and Strømsø (2016) also stated that students' background, such as topic familiarity, "influence[es] their sourcing behavior" (Salmerón, Gila, and Bråten, 2018, p. 25), for this sourcing information assists in predicting and interpreting document content (Barzilai, Tzadok, and Eshet-Alkalai, 2015, in Salmerón, Gila, and Bråten, 2018, p. 25). According to a prior study reported by the NRC (2008), students lack the skills required for engaging in scientific argumentation. They struggle to find relevant data to support claims and provide reliable evidence (McNeill and Krajcik, 2007).

Among the different ethnicities, African Americans scored higher than Latinx students on TU, RC, CK, ASQ, and TAQ, while Latinx students scored higher on ACQ and FAA. African American students scored higher than Latinx students on TF, while Latinx students scored higher on TI. Prior studies shows that only 23% of Latinx and 16% of African American students reach their reading potential, when compared to 76% of students of European descent (Wolf, 2018). However, students of color achieve larger gains in math and reading scores on standardized tests than white students (Shakeel, 2022), according to the 2022 report (Shakeel and Peterson, 2022; Shakeel, 2022). It has been noted that the median rate of progress made by the average African American student is mostly noticeable in elementary school students, but the gap diminishes by high school.

An examination of First Generation and non-First Generation student groups revealed that TU, RC, ACQ, and FAA were lower for the former groups than for the latter (students whose parents had some college and/or had graduated from college); however, on CK, ASQ, and TAQ, as well as on TF, TI, and DRT, First Generation students scored higher. Parental educational attainment and family income were strong predictors of students' achievement among students of different races (HCSPNR, 2022). Historical trends in parental education, income, and ethnicity have been linked to students' achievement (Hanushek, 2020, p. 4).

In terms of Pell Eligibility, TU, RC, ACQ, TAQ, FAA, TI, and DRT scores were lower for Pell Eligible students, but higher when it came to CK, ASQ, and TF. The gap in the achievement between children from high- and low-SES backgrounds has not changed, and according to PISA, the socio-economic status of students has an impact on student performance, in particular in paper-based assessment (OECD, 2015); however, strong readers perform well under both text conditions, paper and digital (OECD, 2018). On a global scale, about 44% of variation in reading performance (mean value) is positively associated with countries of higher per capita GDP (p. 32), but only 8.7% of students in OECD countries are top performers (at level 5-9). Additionally, the OECD study shows an association of about 30% between socio-economic status and reading performance, especially when considering students' self-perception (OECD, 2019). Also, according to *A Half Century of Student Progress Nationwide Report* (Shakeel, 2022), there is a direct relationship between achievement and socio-economic status.

For TU, female students scored higher than male students, African Americans scored higher than Latinx, non-First Generation students scored higher than First Generation, and non-Pell Eligible scored higher than Pell Eligible. On RC, students in the sample fell into the proficient level (low level), and female students scored higher than male students. In comparison to the PISA 2018 countries, girls significantly outperformed boys in reading, by an average of 30 score points across OECD countries, while in the US the gender gap in reading was 24 score points. Also, on RC, African Americans scored higher than Latinx, non-First Generation scored higher than First Generation students, and non-Pell Eligible scored higher than Pell Eligible students.

For CK, students in the sample fell into the advanced-level group; moreover, female scored higher than male students, African Americans scored higher than Latinx, First Generation scored higher than non-First Generation students, and Pell Eligible scored higher than non-Pell Eligible students.

As for ACQ, female scored higher than male students, Latinx scored higher than African Americans, non-First Generation scored higher than First Generation students, and non-Pell Eligible scored higher than Pell Eligible. In terms of ASQ, students in the sample fell into the low-level group, with male students scoring higher than female students, African Americans scoring higher than Latinx, First Generation scoring higher than non-First Generation students, and Pell Eligible scoring higher than non-Pell Eligible students. On TAQ, students in the sample fell into the high-level group: female scored higher than male students, African Americans scored higher than Latinx, First Generation scored higher than non-First Generation students, and non-Pell Eligible scored higher than Pell Eligible students. Finally, in terms of FAA, female scored higher than male students, Latinx scored higher than African Americans, First Generation scored lower than non-First Generation students, and Pell Eligible scored lower than non-Pell Eligible students. Similar trends were observed on a national scale, where Latinx students' performance gains in math and reading are promising when compared to that of white students, although this increase is less noted in reading than in math, because of the language barrier. Latinx students contribute to 78% of English-language-learners in the U.S.

Among the 22 students top performing on TU, the groups that scored the highest (11-14) were females, First-Generation students, African Americans (except for three who were Latinx), and about half of Pell Eligible students. At the highest level, 14, were students who were interested in and somewhat familiar with the topic, as they scored in the medium range on ACQ, but answered FAA question correctly, while scoring at an advanced level on RC and CK. At level 13, most of the students' groups were First Generation, of mixed gender, interested in the topic, and scoring correctly on FAA; however, their ACQ score was at a medium level, their RC score was emerging to proficient, and their CK score was advanced. At level 12 scored mostly female First Generation students who were interested in, but not familiar with the topic; they scored medium to high on ACQ, answered FAA correctly, and their CK score was advanced. At level 11 scored female students, students who were somewhat interested to interested, and somewhat familiar to not familiar; their ACQ score ranged from medium to high, and their CK score was advanced.

Among the 17 bottom performers on TU were students of mixed gender; most were African American, except for one who was Latinx; they were First Generation students, with less than five years of DRT. For example, at level 4 (the lowest level recorded) were students who incorrectly responded to FAA, but their RC score was advanced, and their CK score was proficient; they were non-Pell Eligible students. At level 5 were students who were mostly non-Pell Eligible and First Generation; they reported their DRT as longer than ten years; their ACQ score was medium, their ASQ score low, and their TAQ score high—most of them were African Americans. Finally, at level 6, most students were female, interested in the topic, and somewhat familiar with the topic; they scored in the medium range on ACQ, and high to very high on TAQ, while their RC was proficient, their CK emerging, and their DRT less than five years. Most of them were African Americans.

According to the PISA reports, US teenagers' performance on international PISA tests has been stagnant since 2000. These results indicate that, among fifteen-year-old students, one-fifth read at the level of a ten-year-old, which is troubling because the labor market demands graduates with the cognitive excellence required for computer and STEM jobs. The demographic trends show a wide gap in students' achievement, and if the difference between underserved and privileged students does not narrow, then the mean achievement scores will remain unchanged for US students. Nationwide average performance is at or above the OECD average in reading and science, but below average in mathematics. US students' mean performance in reading has been stable since 2000, in mathematics since 2003, and in science since 2006. Promising results come from the 2018 PISA test, where there is a noticeable increase in performance: in reading, the group of fifteen-year-old students scoring at level 5 or 6 (top performers) increased from about 4% (2009) to 13.5% (2018).

Among gender groups, only ACQ and FAA emerged as statistically significant, and female students performed better than male students on ACQ and FAA. Overall, First Generation and non-Pell Eligible students performed better than non-First Generation or Pell Eligible students, and the results were statistically significant for RC. The ACQ score approached the significance level in the entire sample, per Pell Eligible group. First Generation students performed better than non-First Generation students on CK and TU. Non-Pell Eligible students scored better than Pell Eligible students on RC.

In comparison, the prior findings from OECD (2019) and PISA (2018) show that a noticeable gender gap exists in reading performance, with female students outperforming male students, while in mathematics males did better than females, and in science both male and female students scored about the same. In terms of ratio, for every three male students, only one female student plans to enter a profession in science and technology, such as engineering.

In comparison, across the OECD countries, boys outperformed girls by 5 score points. While girls slightly outperformed boys in science (by 2 score points) on average across OECD countries, according to PISA 2018, US girls and boys performed similarly (OECD, 2019; PISA, 2018).

The article about Earth's magnetism was interesting to only 40% of participants; moreover, only 15% reported being knowledgeable about the topic, even though Earth's magnetisms is one of the most basic topics covered in undergraduate-level geology courses. Our findings are similar to those of King (2012), who explored the most interesting and difficult topics in geology, as reported by undergraduates across three countries: Canada, the UK, and the US. Student participants did not include Earth's magnetism on the list of the most interesting topics in geology. Also, the study by William Boone (1992) evaluating the geology curriculum for non-majors did not even mention Earth's magnetism. Apparently, the topic of Earth's magnetism is not at the top of the list when discussing Earth sciences; thus, perhaps students would score higher on the reading comprehension tasks if the topic were different, more popular, more engaging, or more interesting. Our study shows that those students who had a greater interest in Earth's magnetism and who were more familiar with the topic scored higher on total understanding; thus, these relationships may be worth investigating further in the future.

Earth's magnetism, a topic deeply explored in geophysics, presents a mathematical application to geology; thus, it may appear more challenging and less interesting to students. However, exploring historical magnetism can serve as an example for a concept that cuts across geology and mathematics and therefore helps students gain experience in the application of calculus to an investigation of geological history. This leads to a policy idea requiring K-12 seniors and undergraduates to participate in scientific debates where they would prepare and demonstrate their ability to persuade their audience by using well-prepared scientific arguments. Student debates are not a new idea, since they have long been applied in legal studies; however, in response to the present findings, students in the sciences would also benefit greatly from such events. The following sections will discuss and place in the literature the qualitative interview results.

Interpretation of Results

Reading and scientific literacy. Strategic reading, while emphasized more in social and legal studies, is an important skill for success in the physical sciences as well. The ability to read in a strategic manner helps to break apart difficult and dense texts (see LSAT Extreme; Kaplan, 2009), which students encounter while reading difficult research publications in the sciences. Gaining perspective of the topic allows the reader to calibrate their focus and to separate the leading components. In social studies, the purpose (why) and persuasion (how) that the author communicates with the reader reveals the "silenced" meaning; a reader not just "decodes" the meaning from the text, but also "interprets" the meaning through the leans of personal and cultural experiences (NCTE, 2004 in Stock, P. L., 2005). In the natural and physical sciences, the main goal has previously been thought to consist of concentrating on the specifics of the content, scientific evidence, and scientific justification; however, this is not the case, for discoveries in the "hard science" also require soft skills, such as persuasion and an underlying purpose. As widely known from the history of science, a lack of persuasive skills kept new discoveries in the dark for centuries. It took hundreds of years for scientists to accept Galileo's hypothesis of the heliocentric model; similarly, the idea of plate tectonics proposed before the twentieth century was dismissed until 1960, and in spite of Wagener's discovery of the continental drift in 1912, it took Harry Hess's ability to synthesize prior findings with the new discovery of seafloor spreading, as well as his ability to explain, to persuade the scientific community in 1950 that the continents are, in fact, moving apart (National Geographic, 2023).

In response, the importance of the development of scientific literacy representing students' preparedness for college was examined here. Investigating the results of students' argumentative skills, crucial when reading a difficult scientific text, assist in our understanding of the level of preparedness of the undergraduate student population. The following sections will explain the surveyed strategies, and why they contribute to the students' total understanding.

Reading comprehension (RC) of argumentative writing in the sciences encompasses reading for claim, evidence, structure, and purpose, as well as for "silenced" meaning (quality of persuasion, or author's implications). In effect, the questionnaire tested the students' skills in locating basic information, paraphrases, inferences, and syntheses, and it also asked students to interpret, analyze, and evaluate. The questions that contributed to RC asked about content, including the main idea (claim), justifications (silent meaning), and evidence. The survey also tested logical reasoning, by asking questions about sourcing as expressed through keywords indicating the quality of the scientific evidence, and how the evidence supports the conclusion (reasoning which contributes to persuasion). To recognize evidence as distinct from assumption, students explored the quality of sources (embedded, specific, and general); these results and their explanations will be elaborated on below.

Students' total understanding (TU) combines the variables of reading comprehension (RC) and content knowledge (CK)—that is, TU represents the students' ability to apply formal logic, as well as the ability to recognize representative keywords. The specific question about the conceptual quality of scientific arguments, derived from Sandoval's Model (Sampson and Clark, 2008), asks students to locate in the text these keywords as "causal elements" contributing to justification, and students use it to break down the text during reading. This survey's results focused on content components, such as quality of claim, evidence, and the data represented by sources, their sufficiency, and the overall argument's causal coherence. The initial questionnaire represented the full range of argument components, based on which this study is based.

Assembly Bill 1705. Interviews revealed the students' experiences in respect to changes mandated by the Assembly Bill (AB) 1705, which aims "for the effective core matriculation services of orientation, assessment and placement, counseling, and other education planning services, and academic intervention" (AB 1705, Ch. 926, 2022) in place of remedial courses. This generally seems to be beneficial to students. The trend shows that the number of students enrolling in and completing transfer college-level courses has notably increased. There has been an "increase from 49 to 67 percent of the English and 26 to 50 percent of the math annual completion" (AB 1705, 2022). Without needing to take too many unrelated courses, students focus on the courses about which they are passionate and report exploring university resources more than ever, while enrolling only in the required classes leading towards graduation in their major. The interview responses assisted the researcher in analyzing the narratives through the lens of AB 1750, thus assessing the students' satisfaction as they are navigating college. Looking at students' satisfaction concerning their college experience, timely graduation in a desired major and within a reasonable budget, this study's qualitative analysis focused on the concepts emphasized in AB 1705, complemented by the components of basic human psychological needs as per the Self-Determination Theory (Ryan and Daci, 1985). The interviewed students were of various ethnicities—that is, Asian, African Americans, White and Latinx—but the majority consisted of Latinx. A few students talked about selecting this particular college because of the availability of specific resources for underrepresented groups of students, such as Latinx, African American, or LGBTQ. Students also indicated that the university offered a wide range of resources, as expected by the mandate of the new AB 1705 (AB 1705, Ch. 926, 2022), which aims at effective core matriculation and the replacement of remedial courses with more resources (for example, tutoring, counseling, writing centers, sports clubs, mindfulness centers, social clubs, study groups, or DSS). One student even stated that the number of resources was overwhelming. Prior research has shown that, among students entering college, those from underrepresented groups were more likely to be redirected to pre-transfer courses than other students (AB 1750, 2022). Now, AB 1705 "prohibits […] from recommending or requiring students to enroll in pretransfer level English or mathematics coursework, except under specified circumstances" (AB 1705, 2022). According to this mandate, as of July 1, 2023, the Office of the Chancellor of California Community Colleges will make students' graduation statistics (including historical trends, trends since 2015, progression and completion of transfer-level math and English courses by ethnicity groups) available on its website, for the purpose of transparency and tracking students' timely course completion. Starting in January 2024, in addition to keeping regular updates on the Chancellor's dashboard, the website presenting the statistical data will serve as evidence for legislative committees (AB 1705, 2022). According to Section 1 of the Legislative Information of AB 1705, Chapter 926, Amendment to Education Code related to Community Colleges, community colleges are required to place students who enter college in transfer courses, not in remedial courses. Once students access transfer-level English and mathematics courses, they will complete them within one year. Also, students will only need to take these courses if they fulfill the needs for their intended major; otherwise, students are not required to take these courses, unless the college can specify the benefit of this coursework to the students' major. During the interviews, one student expressed a negative opinion about taking pre-transfer courses during the summer. That student also said that the university would be better off to offer more courses required for the certificate in their major, or for the BA or MA degree. The student explained that, once these courses for art design and film animation were in place, many students would come back. The lack of good math skills prevented another student from completing a degree in civil engineering. Apparently, the difficulty of grasping math skills continues to be a formidable obstacle. AB 1705 states that "(r)esearch has consistently demonstrated that when students are placed and enrolled directly into transfer-level written communication and quantitative reasoning courses, completion of transfer-level coursework in those disciplines is expedited and persistent opportunity gaps in completion of those courses are diminished" (AB 1705, 2022). AB 1705 is a continuum of the previous Bill 705 (2017-18, 2022), which brought positive results in terms of the number of students enrolling in and completing English and math courses within one year. In effect, college enrollment in transfer courses increased for all population groups, with underrepresented groups achieving the largest gains in their completion (they are surpassing white students, although the inequality gap still exists), according to the 2015-19 data included in AB 1705. Apparently, students who are not encouraged to take pre-transfer-level courses graduate sooner. Studies show that underrepresented groups are more likely than white students to enroll in remedial courses. With AB 1705 in place, students' expressed enjoyment in their course of study, reported not being overwhelmed by the number of required classes, and were happy with the reduced number of units and reduced cost for college, which is the goal of AB 1705.

Limitations

This study has limitations that derive from the sampling of students from Earth Sciences classes at one public institution of higher education; thus, the results may not apply to other science classes, nor to other institutions. Some variables had substantial missing data due students missing or not responding to all survey questions. Also, the opinions came from students who were still enrolled at the university, while students who dropped out might think otherwise.

Another limitation worth mentioning is the time frame and the limited 30-minute time block for the students to complete their surveys. On one hand, this allowed for a timely completion of the surveys and interviews, while on the other hand only a limited number of questions could be asked.

Further Research

In this study, we undertook an assessment of students' preparedness for college by examining their performance in the reading comprehension of scientific texts, including their evaluation of the quality of argumentative science writing. The study involved reading and locating significant components of scientific argument, in either paper or digital texts. These quality components consisted of evidence, justification, inferences, and sources, as they have been described in previous studies—such as Sandoval's and Britt's Models. Not all components were assessed at this time; for example, justification was excluded and, therefore, a future study could investigate justification or persuasion strategies. Testing how students assess justification and persuasion quality in science writing may be done in various ways: either in a study similar to the present one, through multiple choice questions, or through students' narrative writing, where participants would, in addition to locating information in the text (argument components), form arguments by writing inferences (intra-text inferences), synthesizing from multiple documents (intertext inferences), and sourcing the evidence.

This study assessed the preparedness of students in Earth Science classes for college based on their own evaluation of the quality of scientific argumentative writing under digital and paper conditions. To answer the research questions, specific independent variables (Gender, Ethnicity, Pell Eligibility, First Generation status, Topic Familiarity, Topic Interest, and digital reading time) were collected from primary and secondary sources. The impact of these variables was explored through five sets of strategies: (1) reading comprehension (RC), (2) Content knowledge (CK), (3) Argument Content Quality (ACQ), (4) Argument Sourcing Quality ASQ, and (5) Total Argument Quality (TAQ) and Final Argument Assessment (FAA). A follow-up study might focus only on Final Argument Assessment, where all the components are evaluated at once.

The undergraduate students' data characterize the demographic characteristics representative of the population of a Southern Californian Institution of Higher Education, as they were collected from lower-level Earth Science classes in a four-year institution of higher education in that region. Future analyses could also look into other student groups from different science classes, and perhaps taught by different teachers, at similar or different higher education institutions. Also, a future analysis may include and annotate the group of students needing extended time.

Participants in this study were surveyed based on scientific texts categorized as difficult, since these texts came from a published scientific journal; another study could include other subjects and levels of difficulty. The instrument originated from previously published articles, noteworthy among them the studies by Sandoval (2005, 2010), Sampson and Clark (2008), Britt (1999), and Salmeron (2018). However, there exist many other studies who present reading and science literacy assessments; thus, using different methods—such as read aloud, or text annotations—may contribute to a greater reliability and validity of the current results.

The scientific arguments used for the students' evaluation were created by the researcher; a follow-up study could instead include the students' own written argumentative narratives. These findings indicate that students did not do particularly well in reading comprehension or argument analyses, and perhaps students need to be better prepared for such analyses. This may be accomplished by providing them with a short lesson on this topic, which would also benefit them in their other courses or further advanced, where they are required to research articles, or perform hands-on research in the laboratory, writing, and presenting their results. More attention needs to be placed on student's preparedness, by providing them with more easily accessible resources supporting the quality of their reading and writing. Unfortunately, as we have learned from the interviews, there is a decreasing trend in the amount of required reading and writing at this higher education institution.

Other types of follow-up research may include a longitudinal study, which may also be beneficial since its results could be compared to trends observed in standardized global and national data. Furthermore, we used secondary data for students' background characteristics, and many variables (for example, Pell Eligibility) that are considered a SES qualifier were missing; hence, these could be obtained through a primary sampling of students for more complete demographics data sets. Finally, assessments could also be performed based on different texts, under desirable conditions, and devices, as mentioned in prior studies: printed text, newspaper, a published article, a book, a digital text on a computer versus a tablet, or mixed media with audio text.

Bibliography

AB 1750 (2022). *Seymour-Campbell Student Success Act of 2012: Matriculation: Assessment*. AB1705 _ California 2021-2022 _ Seymour...matriculation_ assessment. _ TrackBill.pdf

Aharony, N. & Bar-Ilan, J. (2018). Students' academic reading preferences: An exploratory study. *Journal of Librarianship and Information Science,* 50(1), 3–13 https://doi.org/10.1177/0961000616656044.

AMNH. What is a theory? https://www.amnh.org/exhibitions/darwin/evolution-today/what-is-a-theory

Angielczyk, K. (2017). What do we mean by "Theory" in Science. *Field Museum. Retrieved June*, *7*, 2022.

Avvisati, F. (2020). The measure of socio-economic status in PISA: A review and some suggested improvements. *Large-Scale Assessments in Education,* 8(1), 1-37.

Barrio-Cantalejo, I. M., Simón-Lorda, P., Carmona-Rega, I., Escalona-Labella, I., Molina Ruiz, A., & Ayudarte Larios, M. (2008). Opinión de los pacientes sobre la legibilidad de los folletos de educación para la salud. *Index de Enfermería*, *17*(4), 236-240.

Baron, N. (2016). Why digital reading is no substitute for print. *New Republic.* https://newrepublic.com/article/135326/digital-reading-no-substitute-print

Bartlett, F. C. (1932). *Remembering*. Cambridge University Press.

Bucciarelli, L., Drew, D.E. (2015). On MOOCs. The Bridge, National Academy of Engineering, Volume 45, Issue 3, September 23, 2015.

Behr, R. (n. d.). *Scientific Evidence in Criminal Cases*. Ralphbehr.net. https://www.ralphbehr.net/scientific-evidence-in-criminal-cases.html

Belland, B. R., Gu, J., Kim, N. J., & Turner, D. J. (2016). An ethnomethodological perspective on how middle school students addressed a water quality problem. *Educational Technology Research and Development, 64*(6), 1135-1161.

Boe, E. E., & Shin, S. (2005). *Is the United States Really Losing the International Horse Race in Academic Achievement?* University of Pennsylvania. https://repository.upenn.edu/gse_pubs/5

Boone, William J., (1992). Evaluating a Geology Curriculum for Non-majors. *ERIC*.

Breakstone, J., McGrew, S., Smith, M., Ortega, T. & Wineburg, S. (2018). Why we need a new approach to teaching digital literacy. *PHI DELTA KAPPAN* 99, 27–32.

Britt, M. A., Perfetti, C. A., Sandak, R., & Rouet, J. F. (1999). Content integration and source separation in learning from multiple texts. *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso*, 209-233.

Barzilai, S., Tzadok, E., & Eshet-Alkalai, Y. (2015). Sourcing while reading divergent expert accounts: Pathways from views of knowing to written argumentation. *Instructional Science*, 43, 737-766.

Bråten, I., Strømsø, H. I., Salmerón, L. (2011). Trust and mistrust when students read multiple information sources about climate change. *Learning and Instruction*, V. 21, I. 2. P. 180-192. ISSN 0959-4752: https://doi.org/10.1016/j.learninstruc.2010.02.002.

Camina, E., & Güell, F. (2017). The neuroanatomical, neurophysiological, and psychological basis of memory: Current models and their origins. *Frontiers in pharmacology*, 8, 438

CSUDH (2020). https://www.csudh.edu/Assets/csudh-sites/ucm/docs/2022-2023-campus-facts.pdf

College Factual (2023). https://www.csudh.edu/Assets/csudh-sites/ucm/docs/2022-2023-campus-facts.pdf

Clark, D., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research on Science Teaching*, 45(3), 293-321.

Clinton, V. (2019). Reading from paper compared to screens: A systematic review and meta-analysis. *Journal of Research in Reading.* 42(2), 288 – 325. https://doi.org/10.1111/1467-9817.12269

Creswell, J. W. (2016). *30 Essential Skills for the Qualitative Researcher*. Sage.

Deci, E. L. & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Plenum.

Deci, E. L., & Ryan, R. M. (2013). *Intrinsic Motivation and Self-Determination in Human Behavior*. Cham: Springer.

Delgado, P., & Salmerón, L. (2022). Cognitive effort in text processing and reading comprehension in print and on tablet: an eye-tracking study. *Discourse Processes*, 59(4), 237-274.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science education*, 84(3), 287-312.

Dündar, H. & Akçayır, M. (2017). Tablet vs. paper: The effect on learners' reading performance. *International Electronic Journal of Elementary Education*, 4(3), 441–450 Retrieved from http://iejee.com/index.php/IEJEE/article/download/188/184.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of research in education*, 32(1), 268-291.

Earl, S. R., Taylor, I. M., Meijen, C., & Passfield, L. (2017). Autonomy and competence frustration in young adolescent classrooms: Different associations with active and passive disengagement. *Learning and Instruction*, *49*, 32-40.

Encyclopedia Britannica, 2012. Figure 2. Paleomagnetism and magnetic reversals preserved in rocks.

Erduran, S., & Osborne, J. (2005). Developing arguments. In S. Alsop, L. Bencze, & E. Pedretti (Eds.), *Analyzing Exemplary Science Teaching: Theoretical Lenses and a Spectrum of Possibilities for Practice* (pp. 106-115). Open University Press.

Fry, R., Kennedy, B., and Funk, C. (2021). STEM Jobs see uneven progress in increasing gender, racial and ethnic diversity. *Pew Research Center*, 1-28.

Giannelli, P. C. (1993). Junk science: The criminal cases. *Journal of Criminal Law & Criminology, 84*, 105.

Hanushek, E., Petrson, P., Talpey, L., & Woessmann, L. (2020). Long-run trends in the U.S. SES-Achievement Gap. *IZA Discussion Papers No. 12971; p. w26764*. Institute of Labor Economics (IZA).

Hermena, E. W., et al. (2017). Reading Rate and Comprehension for Text Presented on Tablet and Paper: Evidence from Arabic. 8 (February), 1–7. https://doi.org/10.3389/fpsyg.2017.00257

Highlights of U.S. PISA (2018) Results Web Report (NCES 2020-166 and NCES 2020-072). U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. Available at https://nces.ed.gov/surveys/pisa/pisa2018/index.asp.

Higgins, J. A., & Johnson, M. K. (2012). Some thoughts on the interaction between perception and reflection. From perception to consciousness: Searching with Anne Treisman, 390-397.

Hoffman, K. (1988). Ancient Magnetic Reversals: Clues to the Geodynamo. *Scientific American*, 258(5), 76-83. Retrieved June 14, 2021, from http://www.jstor.org/ stable/24989088. See whole Hoffman's article in Appendixes

Johnson, M. K., Raye, C. L., Mitchell, K. J., & Ankudowich, E. (2012). The cognitive neuroscience of true and false memories. True and false recovered memories: Toward a reconciliation of the debate, 15-52.

Hou, J., Rashid, J. & Lee, K.M. (2017). Cognitive map or medium materiality? Reading on paper and screen. Computers in Human Behavior, 67, 84–94 https://doi.org/10.1016/j.chb.2016.10.014.

Kammerer, Y., Keck, D, Starauschek, E., (2019). Effects of a short classroom intervention on students' identification of contradictions in an Internet forum text: Interaction with reading comprehension skills. *Computers & Education* 138 (2019) 46–56

Kavitha, B., Einstein, A., Sivapathasundharam, B., & Saraswathi, T. (2009). Limitations in forensic odontology. *Journal of Forensic Dental Sciences, 1*(1), 8-10.

Kazanci, Z. (2015). University students' preferences of reading from a printed paper or a digital screen: A longitudinal study. *International Journal of Culture and History 1*(1), 50-53.

Kelly, D. L. (2002). The TIMSS 1995 international benchmarks of mathematics and science achievement: Profiles of world class performance at fourth and eighth grades. Educational Research and Evaluation, 8(1), 41-54.

King, H. (2012). Student difficulties in learning geoscience. Planet, 25(1), 40-47.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163-182.

Knight, A. M., & Grymonpré, K. (2013). Assessing students' arguments: How strong are their justifications? *Science Scope, 36*(9), 51.

Knight, A. M., Alves, C. B., Cannady, M. A., McNeill, K. L., Pearson, P. D., Knight, A. M., ... & Pearson, P. D. (2014). Assessing middle school students' abilities to critique scientific evidence. *Annual Meeting of NARST*, Pittsburg, PA.

Kneupper, C. W. (1978). Teaching argument: An introduction to the Toulmin Model. *College Composition and Communication, 29*(3), 237–241.

Koehler, J. J. (2018). How trial judges should think about forensic science evidence. *Judicature, 102*, 28.

Kong, Y., Seo, Y.S. & Zhai, L. (2018). Comparison of reading performance on screen and on paper: A metanalysis. *Computers & Education*, 123, 138–149 https://doi.org/10.1016/j.compedu.2018.05.005

Kopran, C. A., Bisanz, G. L,, Bisanz, J., & Henderson, J. M. (1997). Assessing literacy in science: Evaluation of scientific news briefs. *Science Education*, 81, 515-532.

Lander et al., (2016) PCAST Releases Report on Forensic Science in Criminal Courts, *White House: Blog* (Sept. 20, 2016, 5:59 AM). https://obamawhitehouse.archives.gov/blog/ 2016/09/20/pcast-releases-report-forensic-science-criminal-courts [https://perma.cc/GZ58- D75M]

Lim, H., & Sireci, S. G. (2017). Linking TIMSS and NAEP assessments to evaluate international trends in achievement. *Education Policy Analysis Archives, 25*, 11-11.

LSAT Extreme. (2009). Kaplan.

Mangen, A., & Van der Weel, A. (2016). The evolution of reading in the age of digitization: an integrative framework for reading research. *Literacy*, 50(3), 116-124.

Margolin, S. J., Driscoll, C., Toland, M. J., & Kegler, J. L. (2013). E-readers, computer screens, or paper: Does reading comprehension change across media platforms? *Applied cognitive psychology*, 27(4), 512-519.

Marshak, S. (2019). *Earth: Portrait of a planet*. WW Norton.

McNeill, J. (2014). Holotype specimens and type citations: General issues. *Taxon*, 63(5), 1112-1113.

McNeill, K. L., & Krajcik, J. (2007). Middle school students' use of appropriate and inappropriate evidence in writing scientific explanations. In M. Lovett & P. Shah (Eds.), *Thinking with Data: The Proceedings of the 33rd Carnegie Symposium on Cognition*. Lawrence Erlbaum Associates, Inc.

Mitchell and Johnson, 2000, Johnson, M. K., Raye, C. L., Mitchell, K. J., & Ankudowich, E. (2012). The cognitive neuroscience of true and false memories. True and false recovered memories: *Toward a reconciliation of the debate*, 15-52.

National Geographic, Educational resources (May 2022), National Geographic Society, National Geographic Society, Jeannie Evers, *Emdash Editing*: https://education.nationalgeographic.org/resource/core

National Oceanic and Atmospheric Administration (2016). *Tsunami Warning System*. Noaa.gov. https://www.noaa.gov/explainers/us-tsunami-warning-system

National Research Council. (2013). *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press. https://doi.org/10.17226/18290.

Norris, S. P. (1995). Learning to live with scientific expertise: Towards a theory of intellectual communalism for guiding science teaching. *Science Education*, 79, 201-217.

Nuttall, C. (1982). *Teaching Reading Skills in a Foreign Language*. Heinemann Educational.

New World Encyclopedia, Nov. 15, 2022. https://www.newworldencyclopedia.org/p/index.php?title=Principle&oldid=1060483

OECD (2021). *21st-Century Readers: Developing Literacy Skills in a Digital World*. OECD Publishing.

OECD (2019[a]). *PISA (2018) Assessment and Analytical Framework, PISA*. Paris: OECD Publishing, Paris. https://doi.org/10.1787/7fda7869-en

OECD (2019[b]). *PISA 2018 Results (Volume III): What School Life Means for Students' Lives*. OECD Publishing.

OECD (2019[c]). *Social Impact Investment 2019 The Impact Imperative for Sustainable Development*. OECD Publishing.

OECD (2018). *Result from Programme for International Student Assessment (PISA)*. OECD Publishing.

OECD (2017), PISA for Development Assessment and Analytical Framework: Reading, Mathematics and Science, Preliminary Version, OECD Publishing, Paris.xt

OECD (2013). *PISA 2015: Draft Science Framework: United Kingdom, France*. OECD Publishing.

Olafsen, A. H., Deci, E. L., & Halvari, H. (2018). Basic psychological needs and work motivation: A longitudinal test of directionality. *Motivation and Emotion*, *42*(2), 178-189.

PCAST (2016). *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. Executive Office of the President President's Council of Advisors on Science and Technology. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf

PISA (2020). *Strategic Direction and Vision for Science*. Oecd.org. https://www.oecd.org/pisa/publications/PISA-2024-Science-Strategic-Vision-Proposal.pdf

PISA (2018) Results (Volume I): *What Students Know and Can Do, PISA*, OECD Publishing, Paris; https://doi.org/10.1787/5f07c754-en).

Reardon, S. F. (2013). The widening income achievement gap. Educational leadership, 70(8), 10-16.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. Whither opportunity, 1(1), 91-116.

Reardon, S. F., & Bischoff, K. (2011). Income inequality and income segregation. American journal of sociology, 116(4), 1092-1153.

Rouet, J.-F., Britt, M. A., Mason, R. A., & Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology*, 88(3), 478–493. https://doi.org/10.1037/0022-0663.88.3.478

Salmerón, L., Gil, L., & Bråten, I. (2018). Effects of reading real versus print-out versions of multiple documents on students' sourcing and integrated understanding. *Contemporary Educational Psychology, 52*, 25-35.

Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, *92*(3), 447-472.

Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *The Journal of the Learning Sciences, 12*(1), 5-51.

Sandoval, W. A., Enyedy, N., Redman, E. H., & Xiao, S. (2019). Organizing a culture of argumentation in elementary science. *International Journal of Science Education, 41*(13), 1848-1869.

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23*(1), 23-55.

Shakeel, M., & Peterson, P. E. (2022). A Half Century of Student Progress Nationwide: First Comprehensive Analysis Finds Broad Gains in Test Scores, with Larger Gains for Students of Color than White Students. *Education Next, 22*(4), 50-58.

Silver, Edward A. (1998). *Improving Mathematics in Middle School: Lessons from TIMSS and Related Research.* Office of Educational Research and Improvement, US Department of Education.

Singer Trakhman, L.M., Alexander, P.A. & Silverman, A.B. (2018). Profiling reading in print and digital mediums. *Learning and Instruction*, 57, 5–17 https://doi.org/10.1016/j.learninstruc.2018.04.001.

Stock, P. L. (2005). 2004 NCTE Presidential Address: Practicing the Scholarship of Teaching: What We Do with the Knowledge We Make. College English, 68(1), 107-121.

Strømsø, H. I., Bråten, I., Britt, M.A. (2010). Reading multiple texts about climate change: The relationship between memory for sources and text comprehension, *Learning and Instruction*, Volume 20, Issue 3, 2010, Pages 192-204, ISSN 0959-4752. https://doi.org/10.1016/j.learninstruc.2009.02.001.

Telenius, M., Yli-Panula, E., Vesterinen, V. M., & Vauras, M. (2020). Argumentation within upper secondary school student groups during virtual science learning: Quality and quantity of spoken argumentation. *Education Sciences, 10*(12), 393.

UCLA (n. d.). *What is the difference between categorical, ordinal and interval variables.* Stats.oarc.ucla.edu. https://stats.oarc.ucla.edu/other/mult-pkg/whatstat/what-is-the-difference-between-categorical-ordinal-and-interval-variables/

UNESCO (ND). Sustainable Development Goals 4 2030. https://www.unesco.org/en/education2030-sdg4

USGS (2015). How the Earth generates Magnetic field?

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. Journal of Research in Science Teaching: *The Official Journal of the National Association for Research in Science Teaching*, 46(6), 716-730.

Wolf, M. (2018'). *Reader, Come Home: The Reading Brain in a Digital World*. HarperCollins.

Wolf, M. (2018"). Skim reading is the new normal. The effect on society is profound. Sat, 25, 09-41.

Wolf, M. (Ed.). (2016). *Dyslexia, Fluency, and the Brain*. Pro Ed.

Wolf, M. (2016). *Tales of Literacy for the 21st Century*. Oxford University Press.

Wolf, M. (2007). *Proust and the Squid: The Story and Science of the Reading Brain*. HarperCollins.

Yore, L., Bisanz, G. L., & Hand, B. M. (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education*, 25, 689-725.

Appendix A – Instruments

To Participants:

The following assignment is concerned with scientific argumentative writing and reading comprehension The scientific text you will read here is about Earth's magnetism.

If you are willing to participate, we ask you to sign a consent form and complete the following:

Read for 10 minutes the article: *Ancient Magnetic Reversals: Clues to the Geodynamo* (Hoffman, 1988). Take 20 minutes to respond to questions related to the Article 1, then you will read three arguments written based on the article and select the best, and finally you will respond to questions related to one of the arguments. There are six sets of multiple-choice questions

We appreciate your participation in this study,

Thank you

The following is Article 1 by Hoffman (1988)

It is a fundamental tenet of geophysics that the Earth's magnetic field can exist in either of two polarity states: a "normal" state, in which north-seeking compass needles point to the geographic north, and a "reverse" state, in which they point to the geographic south (Van Zij, 1960). Magnetic reversals have been recorded in magnetically oriented dipoles in minerals of basaltic rocks forming at the oceanic floor at the divergent plate boundaries (Bernard Brunhes, 1906). "Brunhes was intrigued by the discovery of rocks that were magnetically oriented in the direction opposite to the Earth's field" (Hoffman, 1988, p. 76). Geological evidence shows that periodically the field's polarity reverses, and that these reversals have been taking place at an increasing rate. Evidence also indicates that the field does not reverse instantaneously from one polarity state to another; rather, the process involves a transition period that typically spans a few thousand years (Hoffman 1988).

Although, this much is known, the underlying causes of the reversal phenomenon are not well understood. It is generally accepted that the magnetic field itself is generated by the motion of free electrons in the outer core, a slowly churning mass of molten metal sandwiched between the Earth's mantle (the region of the Earth's interior lying below the crust) and its solid inner core. We can conclude that the temperature and density variation of the liquid core, due to pull of gravity and Earth's rotation, provide the driving forces that generate the field. "Phillip L. McFad of the Bureau of Mineral Resources, Geology and Geophysics in Australia and Ronald T. Merrill of the University of Washington suggest that the triggering process is intimately related to the way the outer core vents its heat into the mantle" (Hoffman, 1988, p. 83). The reversal phenomenon may be triggered when something disturbs the heat circulation pattern and in effect the direction of the magnetic field (Hoffman, 1988).

Several explanations for this phenomenon have been proposed. One is the Theory of Heat Transfer through convection, which explains the process in which the outer core vents its heat into the mantle. For example, such heat transfer could create hotter (rising) or cooler (descending) blobs of material from the inner and outer boundaries of the fluid core, thereby perturbing the main heat-circulation pattern (Hoffman, 1988).

A more controversial alternative proposal is the "asteroid-impact." The theory was "proposed by Richard A. Muller and Donald E. Morris of the Lawrence Berkeley Laboratory" (Hoffman, 1988, p. 83). In this scenario an extended period of cold and darkness results from the impact of an asteroid large enough to send a great cloud of dust into the atmosphere. Following this climatic change, ocean temperatures drop, and the polar ice caps grow, redistributing the Earth's seawater. This redistribution increases the rotational acceleration of the mantle, causing friction and turbulence near the outer core-mantle boundary and initiating a reversal of the magnetic field (Hoffman, 1988, p. 83).

How well do these hypotheses account for such observations as the long-term increase in the frequency of reversal? In support of the asteroid-impact it has been argued that the gradual cooling of the average ocean temperature would enable progressively smaller asteroid impacts (which are known to occur more frequently than larger impacts) to cool the Earth's climate sufficiently to induce ice-cap growth and reversals. The theories that depend on extraterrestrial intervention seems less convincing than theories like the first one, which accounts for the phenomenon solely by means of the thermodynamic state of the outer core and its effect on the mantle" (Muller and Morris in Hoffman, 1988, p. 83).

Assessment of Article 1

1. Which statement regarding the Earth's outer core is best supported in Article 1?

A. Asteroid impacts on Earth contributed to how the heat is vented from the outer core into mantle. (0)

B. Motion of electrons within the metallic fluid in the outer core produces Earth's magnetic field. (1)

C. Friction and turbulence, near the outer core and mantle boundary, are typically caused by asteroid impacts. (0)

2. The author talks about temperature blobs in outer core (paragraph 3) in order to:

A. Present a way in which the venting of heat from the outer core might disturb the heat-circulation pattern within the outer core (1)

B. Provide proof for the proposal that ventilation of heat from the outer core into the mantle triggers polarity reversal (0)

C. Give an example of the way in which heat circulates between the Earth's core and the Earth's exterior (0)

3. Which one is the best supported by information in the passage?

A. Most, but not all, geophysicists agree that the Earth's magnetic field may exist in two distinct polarity states. (0)

B. Changes in the polarity of the Earth's magnetic field have occurred more often in the recent past than in the distant past. (1)

C. Heat transfer would cause reversals of the polarity of the Earth's magnetic field to occur more quickly than would asteroid impact. (0)

4. Which one can be inferred regarding the two proposals discussed in the passage?

A. Since their introduction they have sharply divided the scientific community. (0)

B. Both were formulated in order to explain changes in the frequency of polarity reversal. (0)

C. They aren't the only proposals scientists have put forward to explain the phenomenon of reversals. (1)

5. Is magnetic field weakening or strengthening?

A. weakening (1)

B. strengthening (0)

C. neither (0)

6. What does change in magnetic field indicate?

A. Flow of liquid in the core has changed or weakened (1)

B. Earth is becoming tectonically inactive (0)

C. Neither (0)

7. There have been several magnetic reversals, already?

A. True (1)

B. False (0)

8. When north-seeking compass needles points to the geographic north, it indicates:

A. Reverse polarity (0)

B. Normal polarity (1)

C. Neither (0)

9. In 1906 physicist Bernard Brunhes proposed that Earth's field has reversed:

A. True (1)

B. False (0)

10. What causes formation of Earth's magnetic field lines?

A. Electron flow in the liquid outer core (2)

B. Liquid in outer core convects and spirals due to temp. density variations and Earth rotation (1)

C. Temperature variations in the rock layers result from Earth geothermal gradient (0)

11. Scientists learn about magnetic pole's reversals from paleomagnetism

A. True (1)

B. False (0)

12. South Atlantic anomaly phenomena is due to magnetic field decrease in intensity.

A. True (1)

B. False (0)

13. This paleomagnetic evidence support the claim of magnetic reversals and serve, as:

A. Relevant supporting evidence, (1 revised to 2)

B. Relevant-contradictory evidence (0 revised to 1)

C. Irrelevant evidence (0)

14. This example about paleomagnetism fulfills the following role in support of the claim:

A. Relevant supporting evidence (1 revised to 2)

B. Relevant contradictory evidence (0 revised to 1)

C. Irrelevant evidence (0)

15. Evidence accepted by scientific community, are:

A. Geologic and paleomagnetic records (1)

B. Social and political records (0)

C. Volcanic explosivity index records (0)

16. The evidence accepted by scientific community in support of science are:

A. Moral (0)

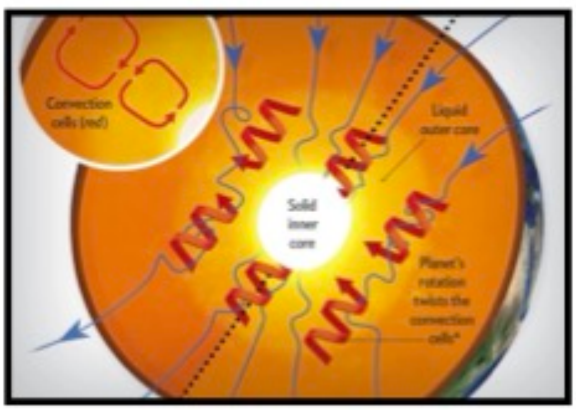
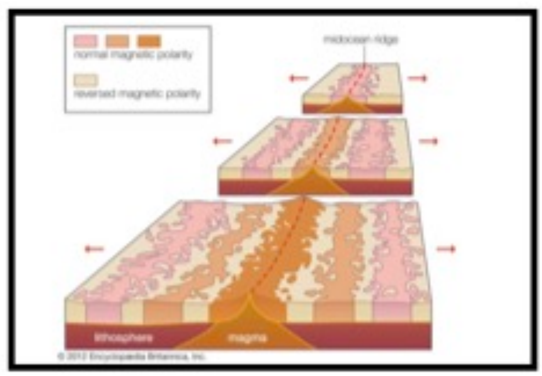
B. Social (0)

C. scientific (1)

D. political (0)

Read Argument A prepared based on Article 1, and respond to questions:

Argument A. According to Hoffman's article, the Earth's magnetic field exists as a result of Earth's outer core liquid convection (Hoffman, 1988) in accordance with Thermodynamics and the Dynamo Theory (education.nationalgeogrphic.org). These convective and spiraling bulbs of molten iron and nickel alloy in the outer core are powered by the temperature variations the Earth's layers and due to Earth's rotation, according to the first article, and shown Figure 1 (p. 11). Hoffman explains that as the bulbs spiral in the outer core they create the flow of liquid and thus the electric flow, (electric current), thus electric and magnetic fields. According to Columbus Law (inverse square law) electrically charged static particles produce electric forces, while moving particles produce electric current and its perpendicular companion magnetic field represented by the magnetic forces along the magnetic field lines (National Geographic, 2022); thus, the electro-magnetic field (2022). The magnetic field lines run from one magnetic pole to another (from N to S). Magnetic poles are not in the same exact positions as the geographic poles. When the magnetic field lines are from N to S (where N is in vicinity of geographic north and magnetic south is in the vicinity of Geographic south) it represents normal magnetism (Hoffman, 1988), but when the magnetic lines are in opposite direction, from S to N (aligned with Geographic South-North direction) than it is called reverse magnetism (1988). As the liquid in outer core flows, it mixes, cools down, slows down and sometime ceases, so does the magnetic field, which weakens. When the liquid core restarts the flow in the same direction the magnetic field strengthens and restarts, but when restarts the flow in opposite direction the magnetic field reverses and the poles switch to opposite sides, Hoffman explains. The direction of magnetic field lines flow might continue in the same direction, disappear, or switch to an opposite direction. The evidence of such reversals is preserved in basaltic rocks and marine sediments, which were discovered during the paleomagnetic studies, Figure 2 (p. 11). The causes of these reversals are not very well known, yet; neither are the causes of changes in the outer liquid core circulation (Hoffman, 1988). Below are figures mentioned in the texts:

<table>
<tr>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>Figure A1. Liquid alloy convection in outer core, and magnetic field lines (Credit: Mark Belan) *</td>
<td>Figure A2. Paleomagnetism and magnetic reversals preserved in rocks (Encyclopedia Britannica, 2012)</td>
</tr>
</table>

17. Argument A contains *inscription (graph or figure):*

A. True (1)

B. False (0)

18. What level of sourcing does Figure 1 represent in Argument A? Choose one that applies:

A. Inclusion (inscription present but there is no reference to it in the text) (0)

B. Pointer (inscription present, it is referenced in the text, but there is no description of it) (0 revised to 1)

C. Description (inscription present, referenced, described, but there is no indication of its role) (1 revised to 4)

D. Assertion (inscription, referenced, described, indicated its role, but no explained why (0 revised to 3)

E. Interpretation (inscription, referenced, described, with its role, indication of why/how (0 revised to 2)

19. Argument A has sufficient # of cited data: (sufficient = all, not just supporting claim

A. Yes (cited data/data support claim and counterclaim) (0)

B. No (cited data/data support only claim) (1)

C. None are present (0)

20. Argument A has a claim

A. Yes, there is a scientific claim or explanation (0)

B. No, there is no claim nor explanation (1)

21. Argument A contains *'General Document Sources':* "article", "text" or "document"

A. No mention of any document use (0)

B. One mention (0 revised to 1)

C. Two mentions or more (1 revised to 2) 3?

22. Argument A contains *'Specific Sources' ex:* accurate author, document, publication

A. No mention of neither author nor document (0)

B. Mention of one author, or one document, or one publication (0 revised to 1)

C. Mention both, author, and document type (0 revised to 2)

D. Mention of all, author, document type and publication type (1 revised to 3)

23. Argument A contains '*Embedded sources' ex: according to IPCC*

A. No mention of embedded sources (0)

B. Mention of one-two embedded sources - insufficient number (0 revised to 1)

C. Mention of three or more embedded sources - sufficient number (1 revised to 2)

24. Argument A has Inferences (source-content link) from one text (Intra-text inferences)

A. No inferences or misconceptions mentioned in the text (no intratext inferences) (0)

B. One idea combined with author’s prior knowledge mentioned (intratext inference) (0 revised to 1)
C. Text has at least one: source-content link (intertext):“ Who said what?” (1 revised to 2)
D. Two ideas combined from one document/author are mentioned in the text (intra-inferences (1 revised to 3)

25. Argument A has Inferences (source-source link), inter-text inferences (cross-texts)

A. Text does not have source-source links, no mention of other author’s ideas (0)
B. Text has one ‘source-source link’ (intertext inference): “author A vs author B” (0 revised to 1)
C. Text has two or more ‘source - source links’ “Author A vs Author B, and Author C vs D” (1 revised to 2)

26. How would you rate the topic interest?

A. Very interesting (1 revised to 3)
B. Interesting (0 revised to 2)
C. Somewhat interesting (0 revised to 1)
D. Not interesting at all (0)

27. How would you rate your knowledge about the topic?

A. Very knowledgeable about the topic, from scientific texts[1] (1 revised to 3)
B. Knowledgeable about the topic the topic (general knowledge from news) (0 revised to 2)
C. Somewhat knowledgeable about the topic (0 revised to 1)
D. Not quite familiar with the topic or unfamiliar (0)

28. How long have you been using digital devise(s) for reading?

A. less than five years (0)
B. between five and ten years (0 revised to 1)
C. more than ten years (1 revised to 2)

29. Select the best scientific Argument from below A, B, C

Hint: Consider quality (3-levels for sourcing from low to high: (1) general, (2) specific, (3) embedded sources]; content [correct and valid claim], and evidence (relevant-supporting, relevant-contradictory, irrelevant evidence], all evidence or some, scientific evidence.)

---

1 Very knowledgeable about the topic refers to having a scientific knowledge from published journals

A. Argument A. (1 revised to 2)

According to Hoffman's article, the Earth's magnetic field exists as a result of Earth's outer core liquid convection (Hoffman, 1988) in accordance with Thermodynamics and the Dynamo Theory (education.nationalgeogrphic.org). These convective and spiraling bulbs of molten iron and nickel alloy in the outer core are powered by the temperature variations the Earth's layers and due to Earth's rotation, according to the first article. Hoffman explains that as the bulbs spiral in the outer core they create the flow of liquid and thus the electric flow, (electric current), thus electric and magnetic fields. According to Columbus Law (inverse square law) electrically charged static particles produce electric forces, while moving particles produce electric current and its perpendicular companion magnetic field represented by the magnetic forces along the magnetic field lines (National Geographic, 2022[2]); thus, the electro-magnetic field (2022). The magnetic field lines run from one magnetic pole to another (from N to S). Magnetic poles are not in the same exact positions as the geographic poles. When the magnetic field lines are from N to S (where N is in vicinity of geographic north and magnetic south is in the vicinity of Geographic south) it represents normal magnetism (Hoffman, 1988), but when the magnetic lines are in opposite direction, from S to N (aligned with Geographic South-North direction) than it is called reverse magnetism (1988). As the liquid in outer core flows, it mixes, cools down, slows down and sometime ceases, so does the magnetic field, which weakens. When the liquid core restarts the flow in the same direction the magnetic field strengthens and restarts, but when restarts the flow in opposite direction the magnetic field reverses and the poles switch to opposite sides, Hoffman explains. The direction of magnetic field lines flow might continue in the same direction, disappear, or switch to an opposite direction. The causes of these reversals are not very well known, yet; neither are the causes of changes in the outer liquid core circulation (Hoffman, 1988).

B. Argument B (0 revised to 1)

According to the first article Earth's magnetic field results from Earth's outer core liquid convection and Earth's rotation. These convective and spiraling bulbs of molten iron and nickel alloy in the outer core are powered by the temperature variation across the Earth's interior and

---

2 National Geographic Society (2022), Educational resources: https://education.nationalgeographic.org/resource/core

due to Earth's rotation, the first article says. As the liquid bulbs rise, spiral, and sink in the outer core, they also control the flow of electrons within the liquid outer core producing electric current, which according to Columbus Law (inverse square law) represents the amount of electric forces between stationary charge particles (direction from plus to minus), and its perpendicular companion, the magnetic forces representing magnetic field, but forming due to moving charge particles, as described in the text. The magnetic field lines run from one magnetic pole to another (from N to S). Magnetic poles are not at the same locations as the geographic poles. When the magnetic field lines flow from the N to S (GN-GS, Geographic North-Geographic south) direction, it represents normal magnetism, but when it flows in opposite direction, after switching to S-N (GS-GN, Geographic South-Geographic north) it is called reverse magnetism. As the liquid outer core flows (rises, spirals and falls), it mixes, cools down, slows down and sometime ceases, in effect the magnetic field weakens. When the liquid core restarts the flow in the same direction the magnetic field strengthens and restarts, but when restarts the flow in the opposite direction the magnetic field reverses and the poles switch to opposite sides. The direction of magnetic field lines flow might continue in the same, disappear, or move in opposite direction [concept-explanation]. The causes of these changes (reversals) are not very well known, yet; neither are the causes of changes in the outer liquid core circulation, the author concludes.

C. Argument C (0)

Earth's magnetic field exists as a result of Earth's outer core liquid convection (Hoffman, 1988). The convective and spiraling bulbs of molten iron and nickel alloy in the outer core are powered by the temperature variations the Earth's layers and due to Earth's rotation [concept explanation - justification], according to the first article. Hoffman explains [embedded source] that as the bulbs spiral in the outer core they create the flow of liquid and thus the electric flow, (electric current), thus electric and magnetic fields [concept]. Electrically charged static particles produce electric forces, while moving particles produce electric current and its perpendicular companion magnetic field represented by the magnetic forces along the magnetic field lines (National Geographic, 2022[3]); thus, the electro-magnetic field (2022). The magnetic field lines run from

3 National Geographic Society (2022), Educational resources: https://education.nationalgeographic.org/resource/core

one magnetic pole to another (from N to S). Magnetic poles are not in the same exact positions as the geographic poles. When the magnetic field lines run from N to S (where N is in vicinity of geographic north and magnetic south is in the vicinity of Geographic south) it represents normal magnetism (Hoffman, 1988), but when the magnetic lines are in opposite direction, from S to N (aligned with Geographic South-North direction) than it is called reverse magnetism (1988). The liquid in outer core flows, it mixes, cools down, slows down and sometime ceases [concept explanation-justification]. The liquid core restarts the flow in the same direction the magnetic field strengthens and restarts, but sometimes restarts the flow in opposite direction, Hoffman writes. The direction of magnetic field lines flow might continue in the same direction, disappear, or switch to an opposite direction. The causes of these reversals are not very well known, yet; neither are the causes of changes in the outer liquid core circulation (Hoffman, 1988).

D. I don't know

Table

Interview Questions

1. Do you feel prepared for the college level courses?
2. What courses do you feel most / least prepared for?
3. Do you feel prepared for college writing?
4. Do you find the amount of reading too excessive or just right?
5. Do you think you have a good foundation for physical and social science classes?
6. How was your transition to the university culture? Do you feel a sense of belonging?
7. Are you involved in the university extracurricular activities and/or study groups?
8. Did you choose your major, or you are undecided. What's your major?
9. Did you choose your major because you feel stronger in this field or just the opposite?
10. What skill set gained in prior education helps your agency towards graduation?
11. What is your experience reading from digital device and paper copy?

Note. Students' from previously surveyed classes (EAR 100 and EAR 101) were approached and requested to participate in interviews. Students' responses were recorded and transcribed.

Appendix B – IRB Documents

Letter from IRB at CGU
Consent Form

E-mail from IRB with granted permission for Exempt status;

Dear Ewa,

Thank you for submitting your research protocol to the IRB at Claremont Graduate University for review. On 12/20/2022, based on the information provided for Protocol #4333 (Scientific Argument Quality Reveals Levels of Science Comprehension in Readers of Multiple Texts), we have certified it as *exempt from IRB supervision* under CGU policy and federal regulations at 45 CFR 46.104(b) (2, 4).

Exempt status means that so long as the study does not vary significantly from the description you have given us, further review in the form of filing annual reports and/or renewal requests is not necessary. Although study termination/closure reports are also not required, they are greatly appreciated. You may specify in relevant study documents, such as consent forms, that CGU human subjects' protection staff members have reviewed the study and determined it to be exempt from IRB supervision. The IRB does not "*approve*" (or disapprove) studies that are exempt, so kindly avoid use of this verb.

If we have approved informed consent/assent forms for your study, please be sure to use the approved versions when obtaining consent from research subjects.

Please note carefully that maintaining exempt status requires that (a) the risks of the study *remain minimal*, that is, as described in the application; (b) that *anonymity or confidentiality* of participants, *or protection* of participants against any higher level of risk due to the internal knowledge or disclosure of identity by the researcher, is maintained as described in the application; (c) that *no deception* is introduced, such as reducing the accuracy or specificity of information about the research protocol that is given to prospective participants; (d) the research *purpose*, *sponsor*, and recruited *study population* remain as described; and (e) the principal investigator (PI) continues and is not replaced.

Changes in *any such features* of the study as described may affect one or more of the conditions of exemption and would very likely warrant a reclassification of the research protocol from exempt status and require additional IRB review. If any such changes are contemplated, please notify the IRB as soon as possible and before the study is begun or changes are implemented. If any events occur during the course of research, such as unexpected adverse consequences to participants, that call into question the features that permitted a determination of exempt status, you must notify the IRB as soon as possible.

Please note that a series of suggestions may also be attached to this email. These are suggestions to develop or improve your research protocol. These suggestions are highly recommended but not required. You do not need to send anything back to the IRB.

If Applicable: Most listservs, websites, and bulletin boards have policies regulating the types of advertisements or solicitations that may be posted, including from whom prior approval must be obtained. Many institutions and even classroom instructors have policies regarding who can solicit potential research participants from among their students, employees, etc., what information must be included in solicitations, and how recruitment notices are distributed or posted. You should familiarize yourself with the policies and approval procedures required of you to recruit for or conduct your study by listservs, websites, institutions, and/or instructors. Approval or exemption by the CGU IRB does not substitute for these approvals or release you from assuring that you have gained appropriate approvals before advertising or conducting your study in such venues.

The IRB may be reached at (909) 607-9406 or via email to irb@cgu.edu. The IRB wishes you well in the conduct of your research project.

Source: IRB Documents, CGU, 2023

Claremont Graduate University

AGREEMENT TO PARTICIPATE IN SCIENTIFIC ARGUMENT QUALITY REVEALS LEVELS OF SCIENCE COMPREHENSION IN READERS OF MULTIPLE TEXTS (IRB # 4333)

You are one of up to 100 students being asked to voluntarily share your assignment as data in a study evaluating reading comprehension and to allow the CSUDH Institutional Research professional, Dr. Elana Olshwang (aolschwang@csudh.edu), to add demographic information about you to this assignment. Once this is done, your name will be removed and the data will be anonymous: People may know that you participated, but no one will know which responses are yours. Allowing this is voluntarily. Whether you agree or disagree will have no effect on your relationship to CSUDH now or in the future. If you agree to share your assignment and then change your mind, you can contact Ewa Burchard at ewa.burchard@cgu.edu, and your assignment will be removed. About 100 students are being asked to participate in this study: Half will read the article online and half will read it in a handout. It will take about 10 minutes to read the article and another 20 minutes to answer questions about it.

STUDY LEADERSHIP: This research project is led by Ewa M. Burchard, PhD Candidate in the School of Educational Studies of the Claremont Graduate University, and supervised by Dr. Drew, Professor of Education, Platt Chair in the Management of Technology.

PURPOSE: The purpose of this study is to learn about students' comprehension of a relatively difficult scientific text (published in peer reviewed journal) presented to them in either of two forms, (1) on paper or (2) on screen.

ELIGIBILITY: To be included in this study, you must be a student in higher education who is 18 years and older; who is a regular student taking Earth Sciences, Physical Sciences or Astronomy classes; and who does not require the time extension to complete the assignment within the specified timeframe.

PARTICIPATION: During the study, you will be asked to complete an assignment within 20-30 minutes; you [the student/participant] will be asked to read for ~10 minutes an article by Hoffman (1988) provided in digital form (in you are in a class assessing digital reading comprehension) or paper form (if you are in a class assessing paper reading comprehension).

Then, you will need to respond the multiple-choice questions within remining ~20 minutes. After completion, you will need to submit your responses through an online platform, as instructed in the assignment. The whole process should not take longer than 30 minutes, and within one class meeting. After, collection of assignment responses, the faculty will pass your data to institutional researcher who would remove your names and IDs and transfer the data to me, the primary researcher of this study. Anonymous responses will be emailed to me through an encrypted email to ewa.burchard@cgu.edu.

RISKS OF PARTICIPATION: There are no risks to allowing your assignment to be data in this study.

BENEFITS OF PARTICIPATION: There will be a scientific benefit to students. We do not expect the study to benefit everyone personally. Although, if you find the reading comprehension of scientific peer reviewed materials useful, it can only benefit you in a positive way for during this participation you would be introduced to components used in scientific argumentative writing, the "moves" authors apply that increase scientific persuasion to convey scientific concepts.

COMPENSATION: Your professor will distribute a $5 gift card to all students who voluntarily agree to share their assignment and participate in this study.

VOLUNTARY PARTICIPATION: Your participation in this study is completely voluntary. You may stop or withdraw from the study at any time without it being held against you. Your decision whether, or not to participate will have no effect on your current or future connection with anyone at CGU or your university or college.

CONFIDENTIALITY: Your individual privacy will be protected in all papers, books, talks, posts, or stories resulting from this study. If we use the data we collect for future research or share it with other researchers, your identity will not be revealed or known. In order to protect the confidentiality of your responses, I will receive your responses after your names and SIDs have been removed. You are one of up to 100 students being asked to voluntarily share your assignment as data in a study evaluating reading comprehension and to allow the CSUDH Institutional Research professional, Dr. Elana Olshwang (aolschwang@csudh.edu), to add demographic information about you to this assignment. Once this is done, your name will be removed and the data will be anonymous: People may know that you participated, but no one will know which responses are yours.

SPONSORSHIP This study is not sponsored.

FURTHER INFORMATION: If you have any questions or would like additional information about this study, please contact Ewa M. Burchard at 310-363-2947, ewa.burchard@cgu.edu. You may also contact Dr. David Drew at 310-713-2946, David.drew@cgu.edu. If you have any ethical concerns about this project or about your rights as a human subject in research, you may contact the CGU IRB at (909) 607-9406 or at irb@cgu.edu. A copy of this form will be given to you if you wish to keep it.

CONSENT: Your signature below means that you understand the information on this form, that someone has answered any and all questions you may have about this study, and you voluntarily agree to participate in it.

Signature of Participant ______________ Date ____________

Printed Name of Participant ____________________________

Source: IRB, CGU, 2023