THE SCIENCE DIAGRAM COMPREHENSION TEST

A Thesis in
Educational Psychology
by
Zachary Myers

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2024
The thesis of Zachary Myers was reviewed and approved by the following:

Peggy Van Meter  
Associate Professor of Education,  
Director of Undergraduate and Graduate Studies  
Thesis Advisor

Jimena Cosso  
Assistant Professor of Education

Matthew McCrudden  
Professor of Education  
Chair of the Graduate Program
ABSTRACT

Diagrams are a common instructional tool used in STEM classrooms. However, students regularly are unable to comprehend these diagrams. This means that students often struggle to learn from visual representations, such as diagrams. While some work has been done to address this concern, it has been hindered by the lack of a measure of diagram comprehension. The present study reports on the development of an instrument that measures diagram comprehension — the Science Diagram Comprehension Test. The results of item-level analysis, exploratory factor analysis, criterion validity analysis, and reliability analysis all indicate that the Science Diagram Comprehension Test successfully evaluates one’s abilities to comprehend the diagrams found in post-secondary science classrooms. By presenting a tool for evaluating diagram comprehension, the present study will enable future research to evaluate the efficacy of instructional interventions aimed at improving student diagram comprehension ability and examine the relationships between diagram comprehension and the ability to learn from diagrams.
# TABLE OF CONTENTS

LIST OF FIGURES .............................................................................................................. vi

LIST OF TABLES .................................................................................................................. vii

Chapter 1  Introduction ........................................................................................................ 1

Diagrams as Visual Representations .................................................................................. 2
  Defining Visual Representations .................................................................................. 2
  The Use of Visuals in Science Education ...................................................................... 4
Comprehension of Visual Representations ......................................................................... 6
  Diagram Comprehension and Learning ....................................................................... 6
  The Comprehension Process ....................................................................................... 7
Challenges with Comprehension ...................................................................................... 10
Supporting Comprehension .............................................................................................. 12
Measuring Diagram Comprehension ................................................................................. 13
  Diagram Comprehension as a Construct .................................................................... 14
Design of the Science Diagram Comprehension Test ......................................................... 16
Prior Data on the Science Diagram Comprehension Test ................................................ 17
  Dataset 1 ...................................................................................................................... 17
  Content Validity .......................................................................................................... 18
Present Study ..................................................................................................................... 21
  The Science Diagram Comprehension Test ................................................................. 21
Research Questions .......................................................................................................... 22
  Item Performance ......................................................................................................... 23
  Validity of the Assessment ......................................................................................... 23
  Scale Properties .......................................................................................................... 26

Chapter 2  Methods .............................................................................................................. 28

Participants ......................................................................................................................... 28
Materials .............................................................................................................................. 30
  The Science Diagram Comprehension Test ................................................................. 30
  Measures of Representational Competence ............................................................... 30
Demographics .................................................................................................................... 31
Procedures ......................................................................................................................... 32
  Recruitment .................................................................................................................. 32
  Collection of Dataset 2 ............................................................................................... 32

Chapter 3  Results ................................................................................................................. 33

Item Properties ................................................................................................................... 33
Exploratory Factor Analysis ............................................................................................. 35
LIST OF FIGURES

Figure 1-1: Example Diagrams ........................................................................................................... 3

Figure 3-1: Item Properties in Dataset 2 .......................................................................................... 35

Figure 3-2: Dataset 1 Scree Plot ..................................................................................................... 37

Figure 3-3: Dataset 2 Scree Plot ..................................................................................................... 37

Figure 3-4: EFA Factor Loadings .................................................................................................... 40

Figure 3-5: Histograms of SDCT Scores ......................................................................................... 44
LIST OF TABLES

Table 1-1: Summary of Research Objectives, Analyses, and Conclusions. ..................22
Table 2-1: Participant Major by Course.................................................................29
Table 2-2: Gender and Race of Participants. ..........................................................29
Table 3-1: Fit Statistics of EFA Models. .................................................................39
Table 3-2: Relationships Between SDCT Scores and Other Variables. .................42
Table 3-3: Scale Properties. ....................................................................................44
Chapter 1

Introduction

Visual representations, such as diagrams, are instructional tools used by the overwhelming majority of science instructors (Linenberger & Holme, 2014, 2015). When used, diagrams can improve learning outcomes and communicate certain information better than language alone (Cromley, Snyder-Hogan, & Luciw-Dubas, 2010). Despite the prevalence of diagrams in the classroom and their learning benefits, students struggle with understanding visual representations (Cohen & Hegarty, 2007, 2012; Kottmeyer, Van Meter, & Cameron, 2020a). While this comprehension problem has been well documented (Cohen & Hegarty, 2007, 2012; Kottmeyer et al., 2020a), there is no measure for assessing students’ abilities to comprehend scientific diagrams across multiple domains. Such a measure is needed, however, in order to evaluate new interventions and instructional techniques designed to improve learning from visual representations.

This paper presents work on the development of the Science Diagram Comprehension Test (SDCT). The SDCT is a multiple-choice test in which respondents see a single diagram and respond to comprehension questions concerning that diagram. The SDCT measures diagram comprehension by testing if one can extract and integrate information from visual representations consistent with the types found in post-secondary natural science classrooms. This study provides reliability and validity evidence to support the interpretation of SDCT scores.
Diagrams as Visual Representations

Defining Visual Representations

At the broadest level, visual representations are objects that stand for something else. Examples of these include maps (Figure 1-1a), line and bar graphs (Figure 1-1b), photographs, anatomical drawings, life-cycle diagrams, and more. The use of visual representations for communicating can be understood as a relationship between the learner, diagram, and content. The diagram is a symbolic representation or sign because it represents another object, or referent – the content (Yakin & Totu, 2014). This sign is then “read” by an interpretant – the learner. There is an intrinsic relationship between the learner, diagram, and content where they all exert influence on one another. The characteristics of an object influence what signs are best for representing the object.

For example, the abstract nature of numbers means that quantitative representations use arbitrary shapes to represent information (e.g., Figure 1-1b) as opposed to geographical maps which rely on the shape of the land they represent (e.g., Figure 1-1a). Furthermore, the characteristics of diagrams can guide one to interpret it in a specific manner such as by making certain portions of it more or less salient (Hegarty, 2014).

The relationship between objects, signs, and interpretation has led to the development of conventions in diagrams. Conventions are parts of the diagram that have specific meanings attached to them to communicate information about the object it represents. These conventions can be arbitrary and may differ by field. For example, Figures 1-1a and 1-1b both provide information about different land regions. However,
Figure 1-1a provides that information based on spatial arrangement whereas Figure 1-1b represents regions using colors. Additionally, Figure 1-1a uses colors to represent different seasons. These differences demonstrate how the same information (location) can be represented in different ways and how the same convention (color) can represent different information. Knowledge of conventions is an essential part of comprehending visual representations: without knowledge of these ways of communication, one is unable to access domain content that is represented using conventions (Hegarty, 2014). Knowledge of conventions is comparable to knowledge of the spelling, grammar, and syntax of a language — it is critical to communication.

**Figure 1-1**

*Example Diagrams*

![Diagram A](image1.png) ![Diagram B](image2.png)

Note. Diagram A is a realistic and moderately complex representation of the fire seasons for different regions of Australia. Diagram B is an abstract and moderately complex representation of thunderstorm frequencies across different regions.

Beyond conventions, diagrams also use analogical relationships to communicate information. Using their visuospatial nature, diagrams have a unique ability to efficiently communicate structural and relational information about the world in a way that language
cannot (Hegarty, 2014). Specifically, diagrams can be made isomorphic – visually and spatially similar – to what they represent which allows them to directly display the relationship between different objects or numbers. For example, a map allows one to interpret the spatial relationship between two locations.

Visual representations vary greatly based on the information they are representing, their reliance on analogical relationships, and their use of conventions. Two primary categories visual representations differ on are abstractness (Ainsworth, 2006; Paivio, 1990) and complexity (Hegarty, 2014). Abstractness is how visually similar a representation is to what it represents and ranges from highly representative to highly abstract (Ainsworth, 2006; Hegarty, 2014). Abstract diagrams do not look similar to what they are representing whereas representative diagrams are visually similar to what they are representing. Complexity refers to how difficult a representation is to “read” which is often conceptualized based on the number of components in the representation and can range from highly simple to highly complex (Kottmeyer et al., 2020a). Abstractness and complexity can vary independently to create representations with different qualities. For example, Figure 1-1a is highly depictive with moderate complexity while Figure 1-1b is abstract but similar in complexity. If one were to include additional locations or values to either figure, the complexity would increase as the number of components increases.

**The Use of Visuals in Science Education**

Visual representations are not a supplemental educational tool in the science classroom. They are central to science education. This is because visual representations
can display objects and ideas that are not directly observable (e.g., numbers, distant locations, microscopic anatomy, etc.), communicate spatial information, and support problem solving (Ainsworth, 2006; Davenport, Yaron, Klahr, & Koedinger, 2008; Hegarty, 2014; Kottmeyer et al., 2020a). Using the visuospatial nature of diagrams can demonstrate how parts of a system interact to make a physical system function (e.g., how the cogs of a machine make contact to rotate each other; Hegarty, 2014). Diagrams can also represent abstract relationships as in Figure 1-1b, which displays frequency values — a non-physical entity. Visual representations can also support problem solving – such as working with fractions – through computational offloading (Ainsworth, 2006). These functions mean that visual representations are a critical tool in science education. They can support teaching biological structures, life cycles, the relationship between physical objects in machines, the structure of chemical compounds, and more.

A natural result of this utility is that diagrams are used extensively in science education. A survey of biochemistry instructors found that more than 90% of instructors utilize diagrams during both instruction and assessment (Linenberger & Holme, 2014). Unsurprisingly, 76% of instructors expect their students to be able to interpret and use a representation to solve a problem (Linenberger & Holme, 2015). Diagrams are also widely prevalent outside of lectures with textbooks allocating a significant amount of space to them (Griffard, 2012; Pozzer-Ardenghi & Roth, 2005). Overall, diagrams serve a central role in STEM education to connect the learner to content in a way that language alone cannot accomplish. Between the wide use of diagrams and their ability to communicate critical information, it is essential that students understand these visuals.
Comprehension of Visual Representations

Diagram Comprehension and Learning

The use of diagrams can improve learning outcomes. This is because diagrams present relational information in a manner that language cannot and promote the construction of mental models (Hegarty, 2014; Mayer, 2014; Schnottz & Bannert, 2003; Paivio, 1990). Specifically, diagrams use visuospatial layouts without a given order to communicate information (Hegarty, 2014; Paivio, 1990). In contrast, language linearly presents information in a clear order (Hegarty, 2014; Paivio, 1990). As a result, diagrams are better suited to communicating certain information than language (e.g., physical structures, relative relations, etc.; Hegarty, 2014). This means that diagrams can be used to enhance learning about these topics.

The benefits of diagrams for learning can be seen throughout multiple studies that have found increases in learning outcomes when learners can comprehend diagrams. For example, Davenport et al. (2008) found that a redesign of a chemistry diagram was associated with significant gains in knowledge tests. Further studies have identified that visual literacy and diagram comprehension skills are major predictors of learning outcomes. A study by Van Meter, Cameron, and Waters (2017) found that a student’s ability to comprehend diagrams had a significant positive relationship with scores on a knowledge post-test taken after learning from a text combined with diagrams. Similar results have been observed with students’ graphical literacy skills predicting their ability to learn and draw inferences from graphs (Shah & Freedman, 2011) and biology diagram comprehension ability predicting learning in biology (Cromley et al., 2013). These
studies demonstrate the importance of being able to understand diagrams for science education.

**The Comprehension Process**

To learn from a visual representation, learners need to comprehend it. During the process of comprehending a visual representation, learners create a mental model of the information (Hegarty, 2014; Mayer, 2014; Paivio, 1990; Schnotz, 2014). Mental models are internal representations of the world that can be constructed through first-hand experience or indirect experiences of the world, such as diagrams (Mayer, 2014; Paivio, 1990). These mental models are critical for learning, communicating information, and problem solving. Typically, the goal of a diagram when used in teaching is to support the development of mental models that can be used to solve certain problems (Linenberger & Holme, 2014). Diagram comprehension is similar to reading comprehension because it is a process supported by numerous specific abilities, influenced by task-specific factors, and has an end product of a mental model that can be further utilized by the learner (Hegarty, 2014; Kottmeyer et al., 2020a).

The construction of mental models through comprehension can be broken down into 1) the selection of relevant material from the diagram, 2) the organization of selected material by building connections across it, and 3) the integration of selected material with prior knowledge (Mayer, 2014). This comprehension process relies on numerous bottom-up and top-down processes (Hegarty, 2014).
Bottom-up processes are driven by design elements and their influence on the comprehension process (Hegarty, 2014). For example, complex representations introduce significant information into the learning environment that the learner must sift through in order to select relevant information (Mayer, 2014). In contrast, simple representations reduce the difficulty of selection which can make the comprehension process easier. Additionally, conventions play a large role in bottom-up comprehension processing. For example, using a zoomed-in section of a larger structure draws attention to that section and promotes the selection of information from that section of the diagram. There are a plethora of factors that will differ across representations and can influence bottom-up processing throughout selection, organization, and integration. Design choices in diagram creation can take advantage of bottom-up processing by using conventions (e.g., highlighting important information, using arrows to depict order, and using color to differentiate certain characteristics of the object).

Top-down factors are equally as important in understanding comprehension (Hegarty, 2014). Prior knowledge can guide the choice of what information is selected from a diagram, the strategies utilized to organize that information, and the ability to integrate that information with other relevant content (Hegarty, 2014). Furthermore, representational familiarity, knowledge of conventions, and a learner’s goal will also influence the comprehension process (Ainsworth, 2006; Hegarty, 2014; Schnotz & Bannert, 2003). These top-down factors will interact with bottom-up factors to create individual differences in comprehension outcomes (Shah & Freedman, 2011).

Take Figure 1-1b for example. Bottom-up factors begin influencing the processing of the representation immediately. The orange bars are more visually salient,
which may draw one’s attention to them first. In contrast, Figure 1-1a has bright colors throughout, meaning that no one part is more salient due to color. Additionally, the use of colors associated with fire in Figure 1-1a may lead students to incorrectly think that the colors represent the severity of fire. Top-down factors, such as one’s goal, can also influence the processing of a diagram. For example, if one views Figure 1-1a intending to understand Australia’s geography, they will focus on the layout of the map—however, they may still find the color distracting despite its irrelevance to the task at hand. In contrast, if they want to understand the fire seasons of different regions they will need to consider both the physical layout and the colors. In this way, bottom-up and top-down factors interact throughout the comprehension process.

There are many ways that top-down and bottom-up factors can support the comprehension of a diagram. For example, the selection of information from a diagram can be supported by making relevant sections visually salient, thus encouraging a learner to attend to the information through bottom-up processing (Davenport et al., 2008; Hegarty, 2014). This can promote the encoding of task-relevant information, a critical step in developing mental models that can be applied to solve problems (Ainsworth, 2006; Hegarty, 2014). This process can be further supported by presenting relevant material close together (e.g., diagrams and their captions) and eliminating any extraneous information (Canham & Hegarty, 2010; Mayer, 2014).

After the selection of relevant information, the learner must now organize this information in a meaningful way (Mayer, 2014). To accomplish this, they must construct a mental model of the information. Mental models are built purposefully and guided by the learner’s sense-making and motivation (e.g., task) for engaging with the material
(Mayer, 2014). From here, the materials are integrated across representation modalities and with prior knowledge.

**Challenges with Comprehension**

The skills that support comprehension are collectively known as representational competence and include identifying important information, understanding conventions, integrating components of a diagram, interpreting meaning, and more (Kragten, Admiraal, & Rijlaarsdam, 2013; Nitz, Ainsworth, Nerdel, & Prechtl, 2014; Rau, 2017). Despite the prevalence of diagrams in the classroom and their centrality in science education, students are rarely taught the skills necessary to comprehend diagrams. For example, while over 90% of biochemistry instructors report using diagrams during instruction and assessment for, less than half of these instructors explicitly teach how to use or draw visual representations (Linenberger & Holme, 2014). Rather, instructors assume that students can comprehend and learn from visual representations without instruction (Pozzer-Ardenghi & Roth, 2005; Schönborn & Anderson, 2009).

This lack of instruction means that students are left to their own devices when it comes to learning from visual representations. Previous work has given students assessments aiming to evaluate diagram comprehension and found that college students struggle with basic comprehension of visual representations. In one study, an assessment testing biology diagram comprehension found an average score of 67% on questions evaluating “low-level comprehension” (Kottmeyer et al., 2020a). A similar study by Cohen and Hegarty (2012) evaluated the spatial reasoning of introductory chemistry
students and found that they had an average score of approximately 66%. These assessments did not require prior content knowledge, but rather emphasized the ability to answer questions about the presented information. Combined, these results demonstrate that students typically struggle to answer basic comprehension questions about visual representations, even when they are answerable without prior content knowledge.

Studies have also found that college students perform significantly worse on tests assessing diagram knowledge than they do on comparable assessments of verbal knowledge (Kottmeyer, Van Meter, & Cameron, 2020b; Van Meter et al., 2017). Learners particularly struggle with comprehending complex and 3D graphics (Cohen & Hegarty, 2007, 2012; Hegarty, 2014). This means that students are expected to learn new content from representations they do not know how to use, a problem termed the representation dilemma (Rau, 2017).

This paints a concerning picture of the use of visual representations in education. Diagrams are widely used and a critical part of both teaching and assessment (Linenberger & Holme, 2014, 2015). This is because diagrams are both essential for communicating certain information (Hegarty, 2014; Mayer, 2014; Schnotz & Bannert, 2003) and can contribute to significantly improved learning outcomes (Davenport et al., 2008). At the same time, students struggle to extract even basic information from a diagram (Hegarty, 2005, 2014; Kottmeyer et al., 2020a) and receive little formal instruction on how to improve their representational competence, creating the representation dilemma (Linenberger & Holme, 2014; Rau, 2017). Given the frequent use of visual representations in the classroom (Linenberger & Holm, 2014, 2015) and the well-documented difficulties many students have in comprehending them (Cohen &
Hegarty, 2007, 2012; Kottmeyer et al., 2020a; Van Meter et al., 2017), it is necessary to develop interventions to support comprehension of visual representations: these interventions require a measure of diagram comprehension to be evaluated.

**Supporting Comprehension**

Comprehension can be supported both by improving the use of representations and improving the skills necessary to comprehend a representation. One external factor that impacts the comprehension of a diagram is the fit between a representation and the task it is being used for. Studies have demonstrated that diagrams that present the same information in different formats support the construction of different mental models (Schnotz & Bannert, 2003). A result of these different mental models is that different diagram formats lead to different ways of thinking about the same information (Ainsworth, 2006; Shah & Freedman, 2011). This means that diagrams that are designed to target a specific type of problem will promote the construction of mental models that allow for more effective engagement with those problems (Davenport et al., 2008; Schnotz & Bannert, 2003). However, supporting comprehension solely by modifying representations is inefficient. There is no guarantee that a representation will always be good nor can a single representation be good in all situations. Additionally, modifying diagrams does nothing to support someone when they need to select a representation, design a representation, or use a representation for a novel task.

A more flexible way to address the representation dilemma is to improve learners’ representational competence directly, such as through classroom instruction. Learner
characteristics play a key role central to diagram comprehension. The construction of mental models from visual representations is influenced by familiarity with the representation type, domain knowledge, convention knowledge, the goal of the learner, strategy use, and much more (Ainsworth, 2006; Hegarty, 2014; Mayer, 2014). This means that several avenues can be used to support diagram comprehension.

It is here, however, that we run into another challenge: measuring diagram comprehension ability. Assessments exist for domain-specific diagram comprehension (e.g., Cromley et al., 2013; Kottmeyer et al., 2020a; Van Meter et al., 2017) or for specific types of representations (e.g., Cohen & Hegarty, 2012; Shah & Freedman, 2011) which indicate that students struggle with this process. However, there is no assessment to evaluate the broad diagram comprehension abilities that are necessary to learn from the variety of diagrams encountered in science courses.

**Measuring Diagram Comprehension**

The lack of a general measure of diagram comprehension is what inspired the development of the Science Diagram Comprehension Test (SDCT). A tool that can be used by educators and researchers to gain insight into a student’s ability to comprehend the type of visuals encountered in science classrooms.

Diagram comprehension is supported by numerous skills; however, the most basic outcome is the ability to extract and connect information from a visual representation without the need to integrate it with domain knowledge. Previous studies have measured this ability in specific domains (e.g., Kottmeyer et al., 2020a) and others the retention of
information from visual learning materials (e.g., McTigue, 2009). The SDCT treats diagram comprehension as a domain-neutral process that is supported by both domain-neutral and domain-specific skills — similar to reading comprehension. Despite this difference from prior work, the results from domain-specific measures can provide insight both into the expected performance and the expected psychometric properties of the SDCT.

Diagram Comprehension as a Construct

Shah and Freedman (2011) found that top-down and bottom-up factors interact when learners are interpreting line and bar graphs. Specifically, content knowledge, graphical knowledge, and diagram format interact to influence one’s understanding (Shah & Freedman, 2011). These results indicate that graphical knowledge is a skill that is distinct from – but related to – content knowledge (Shah & Freedman, 2011). This supports treating diagram comprehension as a general competency distinct from content knowledge – similar to reading comprehension. Nitz et al. (2014) provide evidence that diagram comprehension is a distinct ability as they conclude that the skills that support comprehension (representational competence) develop differently from content knowledge. These studies support the divergent validity of diagram comprehension as its own construct.

Previous work evaluating diagram comprehension has also found multiple factors underlying the construct. A study by Cromley et al. (2013) administered a test of biology diagram comprehension to students. Exploratory factor analysis supported the extraction
of a two-factor model – questions that can be solved in a single step (literal comprehension) and those that require multiple steps (inferential comprehension). This is similar to previous work by McTigue (2009) which did not analyze the factor structure of a measure, but did conceptualize question types as explicit or inferential.

While this prior work on diagram comprehension has produced promising results, these assessments all have notable limitations. First, most only assess one or two domains such as biology and geosciences (e.g., Cohen & Hegarty, 2012; Cromley et al., 2013; Kottmeyer et al., 2020). Second, some assessments only evaluated basic comprehension questions, rather than more complex questions that rely on drawing inferences (e.g., Kottmeyer et al., 2020a). Finally, some only evaluated certain types of diagrams, such as quantitative graphs (Cromley et al., 2010; Shah & Freedman, 2011). Beyond the limitations of assessment design, these prior measures have been developed to study a specific research question. Therefore, the psychometric properties of the assessments have not been a central question. Combined, this means that there is no assessment designed to evaluate comprehension of the broad range of diagrams students encounter in the classroom — let alone one with rigorous psychometric research underlying it.

In sum, the previous work on diagram comprehension supports some important conclusions. First, diagram comprehension appears to exist as its own distinct construct (Cromley et al., 2013; Shah & Freedman, 2011). Second, students typically score poorly on assessments of diagram comprehension (Cohen & Hegarty, 2012; Kottmeyer et al., 2020a). Finally, prior work has conceptualized or identified a two-factor model for diagram comprehension based on literal and inferential comprehension (Cromley et al., 2013; McTigue, 2009). The Science Diagram Comprehension Test addresses the gaps left
by prior comprehension assessments by developing a test that includes a variety of
diagrams covering different levels of abstractness, complexity, and domains while also
asking both literal and inferential questions.

**Design of the Science Diagram Comprehension Test**

The Science Diagram Comprehension Test (SDCT) was developed to address the
need for a measure of diagram comprehension ability. The SDCT includes several
diagrams from multiple subjects and of varying complexity and abstractness. The initial
test included 23 diagrams from major domains of study such as biology, chemistry,
physics, engineering, and included many types of diagrams (e.g., line, bar, schematics,
structures, etc.). These diagrams were selected by a three-person team with expertise in
diagram comprehension and experience with STEM teaching and learning. The diagrams
were selected from a variety of open-source resources and represent those found in a
typical undergraduate science class (see section on Content Validity). The research team
classified diagrams as “simple” or “complex” based on the number of components (e.g.,
labels, zoom levels, etc.) in the diagram (Kottmeyer et al., 2020a). Overall, 11 diagrams
were classified as simple and 12 as complex. These diagrams also vary in how abstract
they are. This provides diagrams that are representative of the complexity, abstractness,
and domain of the diagrams that undergraduate students are likely to encounter.

Each diagram had between one and three questions paired with it, making for a
total of 48 questions. The questions were written as either literal questions or inferential
questions. The literal questions require one to identify information that is explicit in the
representation (e.g., identify a component, read a given value). In contrast, inferential questions require inferencing to elaborate on or connect provided information (e.g., determine the relationship between two components, identify the main idea). The initial SDCT included 24 literal and 24 inferential questions.

The SDCT provides a measure of overall comprehension ability as opposed to a test of specific skills that can support comprehension. The questions in the SDCT were designed to be answered without prior content knowledge. All information necessary to answer the questions was provided by the diagrams. This means that the SDCT evaluates if one can extract and connect information contained within a single diagram. So long as one can comprehend the diagram, they should be able to answer the question correctly.

**Prior Data on the Science Diagram Comprehension Test**

**Dataset 1**

A large exploratory study of undergraduate students \((n = 616)\) was conducted using the original SDCT to evaluate overall reliability, basic validity evidence, and item-level properties. Item-level characteristics (e.g., item-total correlation, difficulty index, and discrimination index) indicated three items should be removed from the SDCT due to a combination of poor item-level characteristics. The remaining 45 items produced an internal consistency estimate of \(\alpha = .89\).

The remaining items were used to analyze validity evidence for the SDCT by comparing performance on the test to related variables. First, students majoring in the
natural sciences scored significantly better than those in the social sciences. Second, more advanced students scored better than more junior students. Finally, Dataset 1 displayed a significant positive correlation between SDCT scores and cumulative GPA. These results provide validity evidence by demonstrating the expected relationship between familiarity with the type of diagrams on the SDCT and SDCT scores as well as the expected relationship between scores and an indicator of academic achievement (i.e., GPA).

**Content Validity**

While the previous data discussed above provides strong evidence that SDCT scores relate to other variables as one would expect, there remains the question of content validity in the SDCT. Specifically, the SDCT assesses the ability of one to comprehend the diagrams found in post-secondary science classrooms. This means that the representations used in the SDCT must be similar to those used in post-secondary science classrooms. To evaluate this, a content validity survey was conducted with a national sample of post-secondary science educators. The sample \( n = 34 \) represented a range of disciplines and institution types. In this survey, participants were asked to evaluate if the diagrams used in the SDCT were “similar to those they would expect undergraduate students to encounter” based on the diagram’s difficulty and abstractness. The content validity survey was developed through a series of three cognitive interviews with post-secondary science educators to ensure that it was clear and concise.

During these cognitive interviews, participants completed the content validity survey while engaging in a think-aloud protocol. This provided insights into confusing
instructions and poorly developed questions. One critical finding during this process was that faculty members often struggled to separate the diagram’s design (e.g., complexity, abstractness, type, etc.) from the content it aimed to communicate. In response to these results, a training was developed to help reviewers separate a diagram’s representational characteristics (such as complexity) from the topic it discusses (see Appendix A). Additional minor changes were made to the content validity survey to improve general clarity based on the cognitive interviews.

In the content validity survey, experts first underwent the training questions to encourage them to focus on the representational characteristics. From here, reviewers were shown each of the 23 diagrams and answered two questions. First, they evaluated the diagram on its similarity to those undergraduate students are likely to encounter on a four-point Likert-type scale. Second, they evaluated the diagram on its similarity to those undergraduate students are likely to understand on a four-point Likert-type scale. After evaluating all the diagrams, participants answered a series of demographic questions. Finally, at the end of the survey participants were asked to provide written feedback for any diagrams that they rated as having poor similarity to those students were likely to encounter. This was done to identify any diagrams that could be edited to be more representative of those used in the classroom. The written feedback was placed at the end of the survey due to concerns that seeing a written response box every time they provided low ratings would discourage participants from evaluating diagrams as having poor similarity.

Responses about the validity of the diagrams were recorded as a 4-point Likert-type scale ranging from “not at all similar” to “extremely similar” to the diagrams.
students are likely to encounter. These responses were then dichotomized to calculate a
content validity index with the lower two options being scored as a decision against the
inclusion of the diagram and the higher two being scored as approval of the diagram. The
item-level content validity index (I-CVI) was calculated as the proportion of reviewers
that endorsed a diagram as appropriate (Polit & Beck, 2006). A threshold of an I-CVI of
.78 was used to determine if a diagram is considered acceptable for use (Polit, Beck &
Owen, 2007). With a large sample \( n = 34 \), the effects of random agreement on the I-
CVI are negligible (Polit et al., 2007). Therefore, no correction for random agreement
was done in order to maintain the interpretability of the I-CVI values.

A total of 14 diagrams were retained with I-CVI values above .78. The expert
reviewers had an intraclass correlation coefficient of .921. From here, a scale content
validity index (S-CVI) was calculated by taking the average I-CVI for the retained
diagrams (Polit & Beck, 2006). The 14 remaining diagrams had an average I-CVI (S-
CVI) of .872. This is slightly below the recommended S-CVI of .9 put forward by Polit et
al. (2007) but higher than the often-used standard of .8 (Davis, 1992). Given the
previously identified challenges experts face in separating representational characteristics
from the specific content of the diagrams and the wide variety of domains included in
both the SDCT and the expert sample, it is not surprising that there was greater-than-
typical disagreement in content validity ratings.

Overall, this prior work provides evidence of acceptable content validity for the
SDCT at both the item- and scale-level. Additionally, the SDCT displayed good item-
level characteristics and reliability.
Present Study

The present study collects additional data on the SDCT (Dataset 2) to provide additional validity and reliability evidence for the SDCT. This additional data was necessary due to changes made in the design of the SDCT in response to the results from Dataset 1 and the expert review.

The Science Diagram Comprehension Test

After reviewing the item-level properties and content validity index values for diagrams, some changes were made to revise the original SDCT. First, the nine diagrams that had I-CVI values below .78 were removed from the revised SDCT. After removing these diagrams, only one item that had poor item-level properties remained. This item was edited to address these issues.

After revisions, the SDCT is composed of 14 diagrams and 28 questions. Nine of the diagrams are classified as simple and five as complex. These diagrams cover numerous fields including biology, chemistry, physics, and geoscience. Two of these diagrams have one question associated with them, ten diagrams have two questions, and two diagrams have three questions. Of the 28 questions, 15 are literal and 13 are inferential. This version of the SDCT can be seen in Appendix B.
Research Questions

The present study takes steps toward developing a rigorous assessment of science diagram comprehension for post-secondary students. As such, the Standards for Educational and Psychological Testing (American Educational Research Association, 2014) were used to guide the development and evaluation of the SDCT. These standards emphasize examining evidence for the quality of assessment items, validity of scores, and reliability of scores. A summary of the research questions, related analyses, and results can be found in Table 1-1.

Table 1-1

Summary of Research Objectives, Analyses, and Conclusions

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Analytic Approach</th>
<th>Major Findings</th>
</tr>
</thead>
</table>
| Item Performance  | • Item-total correlation  
• Difficulty index  
• Discrimination index | • Item 48 should be excluded from the SDCT due to poor performance. |
| Internal Structure | • Exploratory Factor Analysis | • Factor analysis supports a 1-factor model, consistent with the claim that the SDCT measures diagram comprehension as a broad construct. |
| Relationship to Criterion | • Correlations  
  ○ GPA  
  ○ Semesters completed  
  ○ Diagram Knowledge Test  
  ○ Diagram Functions Test  
• Group Comparisons  
  ○ Major (STEM vs Other) | • The SDCT exhibited the relationships one would expect if it was measuring diagram comprehension. |
| Scale Properties  | • Internal Reliability (Cronbach’s Alpha and McDonald’s Omega)  
• Skewness & Kurtosis  
• Average score | • The SDCT exhibited high reliability.  
• The SDCT exhibited acceptable skewness and kurtosis.  
• Average performance was consistent with prior research. |
Item Performance

Do the items in the SDCT display acceptable item-level performance? Broadly speaking, test items aim to differentiate amongst individuals based on a certain characteristic — such as diagram comprehension. There are many statistics available to evaluate if an item does this. These values must be interpreted collectively to determine if items should be removed from the measure. Common metrics include the item-total correlation, difficulty index, and discrimination index of an item.

Ideally, Item-total correlations should be greater than .20 (Nunnally, 1967; Qin, 2006). Difficulty indices should have a wide range (.20 to .80) with some easy and some difficult items (Allen & Yen, 2002). Discrimination indices should be greater than .20 (Aggarwal, 1998). Combined, these values provide insight into how related an item is to the other items in the assessment, how difficult an item is, and how good it is at differentiating amongst participants. Overall, the specific research question regarding the item performance is:

1. Based on item-total correlations, difficulty indices, and discrimination indices, should any items be removed from the assessment?

Validity of the Assessment

Is there evidence to support the validity of the SDCT as a measure of diagram comprehension? Validity refers to the claim that an assessment measures what it purports to measure. It is important to note that tests, broadly speaking, are not valid nor invalid. Rather, they have (or lack) evidence that a given interpretation of assessment scores is
appropriate in a given use-case of an assessment (Messick, 1995). Should the interpretation of scores or the use-case of the assessment change, so too should the evaluation of the validity evidence. Evidence for the content validity of the SDCT’s diagrams has previously been evaluated (as detailed above), which led to revisions of the SDCT. The present study analyzes the SDCT’s internal structure and its relationship to other variables to demonstrate further validity evidence.

The validity of the SDCT can be supported using factor analysis to evaluate its internal structure. Exploratory factor analysis evaluates how assessment questions interrelate. This can support the evaluation of an assessment’s validity by allowing one to compare the behavior of assessment questions to what would be expected based on theory. A one-factor structure would be most consistent with the goal of measuring general diagram comprehension ability. However, prior work on diagram comprehension has supported a two-factor model based on literal and inferential comprehension (Cromley et al., 2013). As such, a two-factor could be considered theoretically consistent. A factor structure based on diagram domain or type would indicate that the SDCT is failing to capture overall diagram comprehension and that the items are primarily assessing content knowledge or familiarity with a specific diagram type. As such, a one-factor model is expected and a two-factor model is considered plausible.

Scores on the SDCT are expected to be related to several other variables. The specific relationships of interest were those between diagram comprehension and representational familiarity (Ainsworth, 2006), academic achievement (Cromley et al., 2010; Kottmeyer et al., 2020a), and representational competence (Nitz et al., 2014; Rau, 2017).
Demographic characteristics were used as indicators of both representational familiarity and academic achievement. First, one would expect students in the natural sciences to obtain higher scores on the SDCT than those in the social sciences due to having content knowledge relevant to the diagrams as well as more experience with science diagrams (i.e., greater representational familiarity). Additionally, more advanced students (in their fourth year) should score better than beginning students (in their first year) for similar reasons. Third, given the fact that diagram comprehension is related to learning outcomes, one would expect that GPA is related to scores on the SDCT (Cromley et al., 2010; Kottmeyer et al., 2020a). Finally, tests of representational competence were also included in this study. These assessments evaluated one’s knowledge of how representations can be used and their ability to identify appropriate diagrams for a task. Representational competence supports comprehension processes (Nitz et al., 2014; Rau, 2017) and should be correlated with scores on the SDCT.

Beyond the presence of these relationships, there should be a pattern in the strength of the relationships which provides an additional source of validity evidence for the SDCT as a measure of overall diagram comprehension. The relationship between SDCT scores and representational competence is expected to be stronger than the relationship between SDCT scores and major, semester-in-school, and GPA. This expectation is because representational competence is more directly related to the specific ability to comprehend diagrams than representational familiarity or GPA. For instance, a number of skills and classes that are not related to diagram comprehension influence one’s GPA. Additionally, major and semester-in-school are indirect indicators of representational familiarity which do not guarantee experience with relevant diagrams.
These anticipated relationships provide opportunities to evaluate the validity of the SDCT based on whether it does — or does not — relate to the other variables as expected.

Overall, the specific research questions on the validity of the SDCT are:

2. Does the SDCT exhibit a factor structure that is consistent with comprehension theories?

3. Does the SDCT correlate as expected to relevant variables such as major, semesters completed, GPA, and representational competence?

**Scale Properties**

Does the assessment display appropriate scale-level properties? Ideally, test scores display a relatively normal distribution with no ceiling or floor effect. This can be evaluated visually, by looking at plots of scores, alongside evaluating the skewness and kurtosis of the scores. Tests should also display good reliability. The reliability of test scores addresses the amount of error variance found in an assessment. A test with high reliability produces consistent scores — i.e., an individual is likely to score similarly across multiple administrations of the assessment.

Cronbach’s alpha is the most widely used coefficient of internal consistency and is based on the intercorrelations of all the test items. However, it is notable that Cronbach’s alpha relies on many assumptions that are rarely met in real data (Sijtsma, 2009) and more robust methods exist. Due to the ubiquity of Cronbach’s alpha, the present study reports it alongside a more robust estimate of reliability as encouraged by Sijtsma (2009). Specifically, McDonald’s omega is used, which utilizes structural
equation modeling to estimate reliability and is interpreted in the same manner as
Cronbach’s alpha (Cho & Kim, 2015). The specific research questions about the scale-
level properties are:

4. Is there evidence that scores on the SDCT are relatively normally distributed
   based on skewness and kurtosis?

5. Is there evidence that the SDCT yields reliable scores as indicated by measures of
   internal consistency (Cronbach’s alpha and McDonald’s omega)?
Chapter 2

Methods

Participants

Participants were undergraduate students recruited from educational psychology, biology, and math classes at a large state school and from economics and math classes at a small liberal arts school (see Tables 2-1 and 2-2). The study was approved by the Institutional Review Board and extra credit was offered to participants for completing the study or an alternative assignment. A total of 592 participants participated in the study. However, a small number (19) did not complete the entire study. Little’s MCAR test indicated that the data was missing completely at random; $\chi^2(117) = 94.2$, $p = .940$. Given that the data was missing completely at random and only a small number of participants did not complete all items, these participants were removed from the dataset giving a final sample size of 573.

An independent samples t-test was conducted to compare the participants from both schools on GPA to evaluate a potential difference in the two populations. A correction for unequal variances was conducted. It was found that students from the R1 institution ($M = 3.35$, $SD = .47$) did significantly differ from students from the liberal arts school ($M = 3.54$, $SD = .40$) in GPA, $t(72.389) = -3.283$, $p < .001$. This may suggest population differences between students from these two samples. However, these differences are extremely small. Additionally, it is unlikely that this is reflective of
differences that would impact the construct of diagram comprehension (e.g., a one-factor model at one institution and a two-factor model at the other) but that is an empirical question beyond the scope of the present study.

**Table 2-1**

*Participant Major by Course*

<table>
<thead>
<tr>
<th>School</th>
<th>Course Topic and Level</th>
<th>Natural Science</th>
<th>Social Science</th>
<th>Humanities</th>
<th>Arts</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1 Institution</td>
<td>Introductory Biology</td>
<td>160</td>
<td>29</td>
<td>1</td>
<td>1</td>
<td>191</td>
</tr>
<tr>
<td></td>
<td>Advanced Biology</td>
<td>52</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Introductory Education</td>
<td>19</td>
<td>119</td>
<td>24</td>
<td>16</td>
<td>178</td>
</tr>
<tr>
<td></td>
<td>Introductory Math</td>
<td>52</td>
<td>16</td>
<td>2</td>
<td>4</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>Did not Indicate</td>
<td>7</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Liberal Arts</td>
<td>Introductory Economics</td>
<td>8</td>
<td>15</td>
<td>3</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Introductory Statistics</td>
<td>12</td>
<td>15</td>
<td>8</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Did not Indicate</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>311</td>
<td>203</td>
<td>38</td>
<td>21</td>
<td>573</td>
</tr>
</tbody>
</table>

**Table 2-2**

*Gender and Race of Participants*

<table>
<thead>
<tr>
<th>Gender</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>174</td>
<td>30.4</td>
</tr>
<tr>
<td>Female</td>
<td>396</td>
<td>69.1</td>
</tr>
<tr>
<td>Nonbinary</td>
<td>2</td>
<td>0.3</td>
</tr>
<tr>
<td>Did not Identify</td>
<td>1</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Race</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>441</td>
<td>77.0</td>
</tr>
<tr>
<td>Black or African American</td>
<td>39</td>
<td>6.8</td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>4</td>
<td>0.7</td>
</tr>
<tr>
<td>Asian</td>
<td>93</td>
<td>16.2</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>5</td>
<td>0.9</td>
</tr>
<tr>
<td>Hispanic or Latino/a</td>
<td>15</td>
<td>2.6</td>
</tr>
<tr>
<td>Middle Eastern or Arab</td>
<td>5</td>
<td>.9</td>
</tr>
</tbody>
</table>

*Note.* Participants were permitted to identify with multiple races, therefore total values are greater than the sample size.
Materials

The Science Diagram Comprehension Test

The SDCT (see Appendix B) was completed by participants. The SDCT is a multiple-choice test which asks questions about various diagrams (see Appendix B for a copy of the SDCT and the previous section on the Science Diagram Comprehension Test for a complete description). Each question is worth one point, meaning that scores can range from zero to 28.

Measures of Representational Competence

The Representations Knowledge Test (RKT) assesses one’s knowledge about how diagrams can be used, what strategies can be used to understand diagrams, and what type of thinking different diagrams can support (see Appendix C). The Representations Functions Test (RFT) assesses one’s ability to evaluate the purpose of a representation or select a representation for a specific task (see Appendix D). The RKT and RFT were included in this study as indicators of representational competence — which supports diagram comprehension (Nitz et al., 2014; Rau, 2017).

Both the RKT and RFT are multiple-choice assessments with each question being worth one point. The RKT is 20 questions, meaning that scores can range from zero to 20. The RFT is eight questions, meaning the scores can range from zero to eight. The RKT and RFT both displayed poor reliability ($\alpha = .608$ and $\alpha = .672$ respectively). While these reliability values are lower than ideal, both measures are currently under
development and attempt to target a wide range of skills that compose representational competence which may be a contributing factor. Lower reliability will reduce the maximum correlation possible between the SDCT and RKT or RFT due to introducing random error (Liu, 1988). Lower reliability values do not increase the risk of identifying a strong relationship when one does not exist. Therefore, they disadvantage the hypothesized correlation between the SDCT and RKT and RFT as opposed to increasing the risk of identifying a false relationship.

**Demographics**

All participants completed a brief demographics survey (see Appendix E) where they provided information about their race, gender identity, and major. They also provided academic information including their major, semesters completed, and GPA. Table 2-1 and Table 2-2 present demographic information. Major and semesters completed were used as indicators of a student’s representational familiarity. This is because advanced students and students in a STEM major are expected to have more academic experience with science diagrams than beginning students and students with non-STEM majors.
Procedures

Recruitment

Participants were recruited from 12 classes at a large R1 institution as well as 2 classes from a small liberal arts college. Both institutions are in the Northeastern United States. Participants were recruited through advertisements posted to their class page along with an announcement made in class. All participants were offered extra credit for completing the study. An alternative assignment was made available for participants who were under the age of 18 or did not want to participate in the study. The list of courses can be seen in Table 2-1 along with a breakdown of participant major. Table 2-2 contains the demographics of the sample.

Collection of Dataset 2

All materials were delivered online via Qualtrics. Participants began by completing informed consent. From here, participants completed the SDCT. The SDCT was presented one question at a time, with questions delivered in a random order. Questions about the same diagram were not shown consecutively. After completing the SDCT, participants completed the representations knowledge test followed by the representations functions test. Finally, participants completed the demographics survey.
Chapter 3

Results

All analyses were done in R version 4.3.1 (R Core Team, 2023) using the Psych package (Revelle, 2023), the Psychometric package (Fletcher, 2022), and the lavaan package (Rosseel, 2012). A summary of research questions and major findings can be found in Table 1-1.

Item Properties

All items except one displayed good performance in Dataset 1. However, it is important to evaluate and ensure that the items continue to perform well in Dataset 2 and assess the redesign of the single item that previously performed poorly. Item-level performance was evaluated by examining the item’s corrected item-total correlation, difficulty index, and discrimination index. Figure 3-1 displays the item properties.

Corrected item-total correlations indicate how an item relates to the total score on a test. This is done by calculating the point-biserial correlation between a test item and total test scores calculated without that item. Values range from -1.0 to 1.0 with a correlation of .30 or greater being seen as good and below .20 as poor (Nunnally, 1967; Qin, 2006). One item displayed an extremely low item-total correlation (item 48). Other items displayed low item-total correlations that were less severe (items 15 and 36).
The difficulty index provides the proportion of students that correctly answered a question, with lower numbers indicating more difficult questions. There are many recommendations on the appropriate range of a difficulty index, however, most agree that items with a difficulty index below .2 or above .8 indicate an item that is too difficult or too easy (Allen & Yen, 2002). One item fell below this range, indicating it was extremely difficult (item 48). Two items fell above this range, indicating that they were somewhat easy (items 25 and 47). However, this is to be expected as the SDCT is designed to be accessible to a typical undergraduate student.

The discrimination index indicates how effective an item is at distinguishing individuals. This is done by taking the lowest and highest scoring 27% of participants, finding how many more individuals in the highest scoring group got the item correct than in the lowest scoring group, and dividing this difference by the total number of participants. Values range from -1.0 to 1.0 with a discrimination index of .2 or greater being seen as acceptable (Aggarwal, 1998). One item fell far below this threshold (item 48) while another (item 3) was slightly below. All other items displayed acceptable discrimination.

All three of these metrics were considered together to determine if an item should be removed from the SDCT. For instance, an item that was rather easy (i.e., had a high difficulty index) but displayed an acceptable item-total correlation and discrimination index should not be removed solely based on one statistic. Item 48 was found to have an extremely low difficulty index, low item-total correlation, and a low discrimination index (see Figure 3-1). The redesigned item (7) performed well. Only item 48 displayed poor performance across all metrics, therefore only this item was removed from the SDCT and
excluded from future analyses. This means that 27 items remained on the SDCT spread across 14 diagrams.

**Figure 3-1**

*Item Properties in Dataset 2*

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Corrected Item-Total Correlation</th>
<th>Difficulty Index</th>
<th>Discrimination Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>48</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*Note:* Dotted lines represent common item-performance cutoffs.

**Exploratory Factor Analysis**

An exploratory factor analysis (EFA) was conducted on both Dataset 1 and Dataset 2. A one factor model was considered most consistent with the goal of measuring overall diagram comprehension ability. However, previous work on diagram comprehension has produced a two-factor model based on inferential and literal questions (Cromley et al., 2013). Therefore, a two-factor model was also considered theoretically viable. Additionally, the items may load onto factors based on the domain of the diagram (e.g., one factor for questions about biology diagrams, one factor for questions about
physics diagrams). This would indicate that the SDCT fails to isolate overall diagram comprehension, and instead measure domain-specific comprehension or content knowledge. The sample sizes for both datasets were large enough to produce good estimates of the true factor structure (Pearson & Mundform, 2010).

Only items that were retained on the SDCT were included in the factor analysis. The KMO measure of sampling adequacy was good for both Dataset 1 \( (MSA = .87) \) and Dataset 2 \( (MSA = .91; \text{Kaiser, 1974}) \). Bartlett’s test of sphericity indicated that the correlation matrices both Datasets was appropriate for factor analysis \( (p < .001 \text{ for both; Dziuban & Shirkey, 1974}) \). Because the items on the SDCT are scored dichotomously, tetrachoric correlations were used with unweighted least squares estimation \( (\text{Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009}) \). Parallel analysis and fit statistics were used to evaluate the number of factors \( (\text{Finch, 2020; Hayton, Allen, & Scarpello, 2004; Yang & Xia, 2015}) \). Parsimax rotation was used to produce the final factor loadings \( (\text{Finch, 2011}) \).

**Factor Structure**

Both factor analyses support a one-factor model. The Kaiser Criterion states that only factors with eigenvalues greater than one should be retained but tends to extract more factors than exist \( (\text{Kaiser, 1960; Yang & Xia, 2015}) \). Parallel analysis tends to be more accurate and retains factors with eigenvalues greater than reference eigenvalues that are simulated from data with no factor structure \( (\text{Yang & Xia, 2015}) \). The parallel analysis used corrected reference-values as detailed by Lubbe (2019). The parallel
analysis of Dataset 1 and Dataset 2 both support a one-factor model. The scree plot for the Dataset 1 EFA can be seen in Figure 3-2 and the Dataset 2 EFA in Figure 3-3. These results provide strong evidence for a one-factor model of the SDCT and are further supported by model fit statistics.

**Figure 3-2**

*Dataset 1 Scree Plot*

**Figure 3-3**

*Dataset 2 Scree Plot*
Model Fit

The one-factor model displayed good overall model fit in both Dataset 1 and Dataset 2. Fit statistics for the models are summarized in Table 3-1. The one-factor model explained 32% of the variance in scores in Dataset 1 and 30% in Dataset 2. Additional fit statistics, such as the root mean square error of approximation (RMSEA) and standardized root mean squared residual (SRMR) suggest that the model has an acceptable fit (see Table 3-1). Common guidelines suggest that the RMSEA should be below .06 and the SRMR below .08 (Hu & Bentler, 1999). Based on Hu & Bentler’s (1999) guidelines, the SDCT displays acceptable fit (see Table 3-1). These results indicate that while the one-factor model fits the data well, there is still a notable amount of variance that cannot be accounted for with this factor model.

This is consistent with the theory that diagram comprehension is an overarching ability which is supported by multiple skills that interact with one another (e.g., conventions knowledge and content knowledge). A one factor model does not account for each of these abilities individually, but rather models the product of their interaction: the extraction of information from a diagram. A significant amount of variance is explained by modeling a single factor that may represent a composite of more foundational abilities. While more variance might be explained by modeling each ability that contributes to diagram comprehension, this would require a significantly longer assessment.
Table 3-1

*Fit Statistics of EFA Models*

<table>
<thead>
<tr>
<th>Factors Extracted</th>
<th>Dataset 1</th>
<th>Dataset 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Variance Explained</td>
<td>RMSEA</td>
</tr>
<tr>
<td>One</td>
<td>.32</td>
<td>.050</td>
</tr>
<tr>
<td>Two</td>
<td>.36</td>
<td>.036</td>
</tr>
<tr>
<td>Change</td>
<td>+.04</td>
<td>-.014</td>
</tr>
</tbody>
</table>

*Note.* RMSEA = Root Mean Square Error of Approximation (lower is better), SRMR = Standardized Root Mean Square Residual (lower is better)

The fit of a two-factor model was also evaluated in both datasets using a Parsimax rotation to evaluate a potential alternative model (Finch, 2011; 2020). Comparing the RMSEA between the one-factor and two-factor models can help inform the appropriateness of adding a second factor. Simulation studies indicate that when adding an additional factor does not reduce the RMSEA by at least .015, then that factor should not be extracted (Finch, 2020). Adding a second factor did not decrease the RMSEA by enough to justify adding a second factor in either dataset (see Table 3-1).

When added, the second factor only had a small number of items loading onto it. The items loading onto the second factor were inconsistent across the two datasets (i.e., the items composing the two factors were not the same). Additionally, many items displayed mild cross-loadings and inconsistent loading directions between the two datasets (e.g., positive in one dataset and negative in the other). The items loading onto the second factor also tended to be more difficult, indicating that the two-factor model may have been splitting items based on their difficulty rather than a relationship with a
truly unique latent trait (Yang & Xia, 2015). This is particularly worth considering as there was not a consistent and theoretically justifiable pattern of items loading onto the second factor (e.g., inferential items, physics items, etc.).

Collectively, the two-factor model displays an improvement in fit and variance explained (see Table 3-1). However, this improvement is relatively small and overshadowed by, 1) the number of cross-loadings, 2) a small number of items loading onto the second factor, 3) inconsistency in which items load onto the second factor, and 4) the lack of a theoretically sound explanation for the observed patterns.

The factor loadings for both one-factor EFA models can be seen in Figure 3-4. All items loaded acceptably onto the single factor. Item 7 was excluded from Dataset 1’s analysis due to poor design that was revised before the collection of Dataset 2.

**Figure 3-4**

*EFA Factor Loadings*

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Dataset 1</th>
<th>Dataset 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Dotted line represents a factor loading of .32.
The average factor loading was .54 in Dataset 1 and .53 in Dataset 2. A common rule of thumb suggests factor loadings should be above .32 because this equates to approximately 10% of the variance explained (Costello & Osborne, 2005; Tabachnick & Fidell, 2001). Three items (15, 32, and 36) showed factor loadings below .32 in both datasets. However, all these items still produced factor loadings above .2, the point at which loadings are often considered negligible (Matsunaga, 2010). Overall, there are several different guidelines for interpreting factor loadings that all must be considered within the context of theory (Matsunaga, 2010). These items were retained because they did not display consistent issues in their item-level properties and are theoretically appropriate. Overall, the SDCT displayed acceptably strong factor loadings.

**Criterion Validity**

There are participant characteristics that one would expect to be related to participant scores. Specifically major, semester standing, and GPA. As discussed with the research questions, individuals majoring in the natural sciences and those with more semesters completed are expected to have higher scores due to increased representational familiarity (Ainsworth, 2006). A positive correlation was also expected with GPA due to the relevance of diagram comprehension for academic achievement (Ainsworth, 2006; Hegarty, 2014; Kottmeyer et al., 2020a; Linenberger & Holme, 2014). Finally, a stronger relationship was expected between diagram comprehension scores and measures of representational competence. The results of all these analyses can be found in Table 3-2.
Table 3-2

*Relationships Between SDCT Scores and Other Variables*

<table>
<thead>
<tr>
<th>SDCT Data</th>
<th>Major</th>
<th>Semesters Completed</th>
<th>GPA</th>
<th>Representations Knowledge Test</th>
<th>Representations Functions Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset 1</td>
<td></td>
<td>$t(614) = 5.93^{***}$</td>
<td>$\rho = .129^{***}$</td>
<td>$r = .147^{**}$</td>
<td>n/a</td>
</tr>
<tr>
<td>Dataset 2</td>
<td></td>
<td>$t(571) = 5.20^{***}$</td>
<td>$\rho = .119^{**}$</td>
<td>$r = .275^{***}$</td>
<td>$r = .555^{***}$</td>
</tr>
</tbody>
</table>

* $p < .05, **p < .01, ***p < .001$

These relationships were evaluated in both Dataset 1 and Dataset 2. This is because the initial criterion validity analyses done in Dataset 1 included diagrams that failed the expert content review. Therefore, these relationships should be reevaluated without those items included.

There was a significant difference in SDCT scores based on having a major in the sciences in both Dataset 1 and Dataset 2. An independent samples t-test was used for this analysis because major was dichotomized into ‘natural science major’ and ‘not natural science major’ due to the limited number of students majoring in the humanities or arts. Additionally, there was a significant correlation between semesters completed and SDCT scores in both Dataset 1 and Dataset 2 (see Table 3-2). Spearman’s Rho was used for this analysis due to the limited scale of semesters completed. Finally, there was a significant correlation between GPA and SDCT scores for both Dataset 1 and Dataset 2. A Pearson correlation was used for this analysis because SDCT scores and GPA are both treated as continuous and interval data. The results of these analyses are reported in Table 3-2.
In addition to these demographics, students in Dataset 2 also completed the Representations Knowledge Test (RKT) and Representations Function Test (RFT). One would expect RKT and RFT scores to correlate with SDCT scores because knowledge of strategies for understanding representations, knowledge of representation functions, and diagram comprehension should support one-another. It was found that both the RKT and RFT correlate significantly with scores on the SDCT (see Table 3-2).

These results demonstrate that SDCT scores relate to relevant variables as expected. Additionally, the pattern of the strength of these relationships is also consistent with expectations. Specifically, scores were related to representational familiarity as indicated by major and semesters completed, academic achievement as indicated by GPA, and representational competence as indicated by the RKT and RFT. The correlations between the SDCT and the RKT and RFT were stronger than the correlations between the SDCT and GPA or semester. This indicates that the SDCT has a stronger relationship with more proximal measures (i.e., measures of representational competence). Collectively, this pattern suggests that scores on the SDCT represent diagram comprehension as an ability applicable across many academic settings that is supported by various other skills.

Scale Properties

The scale properties are summarized in Table 3-3 and scores are plotted in Figure 3-5. The statistics were calculated for both Dataset 1 and Dataset 2 because the initial analysis of Dataset 1 included diagrams that failed the expert content validity review.
The average scores were 68.1% and 64.2% for Dataset 1 and Dataset 2 respectively (see Table 3-3). In Dataset 1, the SDCT displayed mild negative skew and negligible kurtosis. In Dataset 2, the SDCT displayed mild negative skew and kurtosis. This suggests a distribution of scores with only mild deviations from normality (see Table 3-3 and Figure 3-5).

Table 3-3

Scale Properties

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Average Score</th>
<th>Stnd. Dev.</th>
<th>Reliability (α)</th>
<th>Reliability (ω)</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset 1</td>
<td>68.1% (of 26)</td>
<td>4.99</td>
<td>.84</td>
<td>.85</td>
<td>-.80</td>
<td>-.05</td>
</tr>
<tr>
<td>Dataset 2</td>
<td>64.2% (of 27)</td>
<td>5.33</td>
<td>.84</td>
<td>.85</td>
<td>-.52</td>
<td>-.57</td>
</tr>
</tbody>
</table>

Figure 3-5

Histogram of SDCT Scores
Internal reliability was evaluated using both Cronbach’s alpha and McDonald’s omega. This follows the recommendations of Sijtsma (2009) which suggests reporting Cronbach’s alpha due to its prevalence alongside more robust estimates of reliability, such as McDonald’s omega (Cho & Kim, 2015). Cronbach’s alpha and McDonald’s Omega are both interpreted in the same manner and can range between 0 and 1 with higher values indicating less error variance in the measure (Cho & Kim, 2015). There are numerous recommendations regarding appropriate values for reliability (see Taber, 2018). Typically, values above .70 are seen as acceptable for new measures and values above .80 considered acceptable for use in research (Nunnally, 1967). The SDCT displayed good reliability in both Dataset 1 ($\alpha = .84$, $\omega = .85$) and Dataset 2 ($\alpha = .84$, $\omega = .85$).
Chapter 4

Discussion

Diagrams and visual representations are abundant in science education (Linenberger & Holme, 2014, 2015). Despite this, students regularly struggle to comprehend diagrams (Cohen & Hegarty, 2007, 2012; Kottmeyer et al., 2020a). While work has been done to improve the design of diagrams and how they are used (e.g., Ainsworth, 2006; Davenport et al., 2008; Mayer, 2014), little work has been done to directly improve student diagram comprehension let alone examine how a student’s comprehension ability influences their use of, and what they learn from, these representations. A major contributor to this is the lack of a robust measure of diagram comprehension that can be applied across domains. The present study addresses this need by evaluating a new measure – the Science Diagram Comprehension Test (SDCT) – to determine the appropriateness of using it to examine diagram comprehension ability as an individual difference in research.

The results and conclusions of this study are summarized in Table 1-1. After revisions, the 27 items retained on the SDCT demonstrated appropriate item performance statistics. The final design of the SDCT included 27 questions paired with 14 diagrams. Nine of the diagrams are considered simple while five are complex. Of the 27 questions, 15 were literal and 12 inferential. The results of EFA suggest that scores on the SDCT should be interpreted as a single total score which represents one’s diagram comprehension ability across multiple domains and diagram types. This interpretation is
further supported by the statistically significant relationships between SDCT scores and relevant variables (e.g., academic achievement, representational familiarity, and representational competence). Finally, the good reliability and relatively normal distribution of scores indicate that the SDCT is appropriate for use in research.

**Performance on the SDCT**

Many items showed higher difficulty indices (suggesting they are easy), but this is not surprising. The questions in the SDCT are designed to be answered solely with the information provided in the diagrams and the diagrams are designed to be representative of those that students encounter in typical science courses. This means that, assuming one can comprehend the diagrams, the questions should be extremely easy. As such, one would expect most undergraduate students to perform well on the SDCT.

Despite the presence of some ‘easier’ questions and the design of the materials to be accessible, scores were low. The average score on the SDCT was 68.1% for Dataset 1 and 64.2% for Dataset 2. From a psychometric perspective, these scores may not seem particularly low. However, the SDCT does not display a ceiling nor a floor effect. No participants scored a zero, only 4 participants achieved a perfect score in Dataset 1, and no participants achieved a perfect score in Dataset 2 (see Figure 3-5).

These results are consistent with prior work demonstrating the struggles students have with comprehending diagrams (Cohen & Hegarty, 2007, 2012; Kottmeyer et al., 2020a), even when asked basic questions that do not require additional background information to answer. The SDCT was designed to require no content knowledge and
does not require individuals to recall the diagram from memory (because it is presented with the questions). As such, the SDCT specifically measures the ability of an individual to extract and connect information from a given visual representation. These average scores suggest that students are unable to do this for approximately one in every three questions. Given that visual representations are in the overwhelming majority of learning materials (Linenberger & Holme, 2014), these findings suggest that, of those materials that include visuals, students are unable to extract basic information from them approximately 33% of the time.

Consistency with Theory

Factor Model

Previous work has supported a two-factor model of comprehension based on literal and inferential comprehension (Cromley et al., 2013). Literal questions simply require identifying information in the diagrams while inferential questions require making connections across pieces of information. Therefore, the SDCT was designed to include both literal and inferential questions. While a one-factor model was consistent with the theory that there is a single diagram comprehension ability, a two-factor model based on literal and inferential comprehension was considered equally plausible. The results indicated a one-factor model of the SDCT. This is consistent with the theory that diagram comprehension can be conceived of as a skill that spans domains, similar to reading comprehension (Kottmeyer et al., 2020a). However, this result is inconsistent
with the findings of Cromley et al. (2013) and the theoretical explicit versus inferential distinction seen in other research (McTigue, 2009).

This prior work has differed from the SDCT in critical ways. For example, McTigue (2009) designed a test of diagram comprehension with explicit and inferential questions. While factor analysis is not reported in this study, the explicit versus inferential distinction is conceptually important. However, McTigue (2009) created this measure to evaluate the effect diagrams have on knowledge recall. This is a significant difference from the design of the SDCT which presents a question alongside a diagram and does not measure recall.

A study by Cromley et al. (2013) measured biology diagram comprehension as part of a larger study evaluating an educational intervention. This study reported that a two-factor model was supported by an EFA of a biology diagram comprehension test. The factors were distinguished based on whether the item required a single step or multiple steps, which the authors classified as literal and inferential respectively (Cromley et al., 2013). There are many differences between the present study and Cromley et al. (2013) which could explain the extraction of different factor models.

First, the study goal was not the development of a comprehension measure, but rather the evaluation of an instructional intervention. As such, all students were expected to learn relevant content before completing the diagram comprehension test and content knowledge was expected to be necessary to answer the diagram comprehension questions (Cromley et al., 2013). The domain-specific nature of the test means that item responses are likely to systematically vary based on biology prior knowledge. This means that a factor which represents content knowledge may have appeared in the study by Cromley
et al. (2013) because content knowledge was an expected requirement for their diagram comprehension measure. However, this leaves the question of why the factors seemed to organize based on inferential or literal questions. One possible explanation for this pattern is that inferential questions rely more heavily on content knowledge. Inferential items require one to connect across pieces of information in the diagram, a process which benefits from the integration of prior knowledge (Hegarty, 2014; Mayer, 2014). Literal items do not ask a student to connect or integrate information and therefore might be less reliant on top-down factors such as prior knowledge (Hegarty, 2014; Mayer, 2014).

Additionally, the questions asked by Cromley et al. (2013) are about interpreting the main idea of a diagram, a process which is influenced by prior knowledge (e.g., Shah & Freedman, 2011). Therefore, the inferential factor may be composed of inferential items because they are more heavily influenced by biology prior knowledge rather than because inferential comprehension and literal comprehension are different abilities.

The SDCT would not display a prior-knowledge factor due to differences in the design. First, the SDCT includes diagrams that cover a wide range of domains. While prior knowledge can support one’s ability to answer questions on the SDCT, it is not systematic across the items because of how many topics are included. For example, the content knowledge relevant to an anatomy diagram is not the same as the prior knowledge relevant to a geological diagram. Second, the SDCT is designed such that questions can be answered solely with information presented in the diagram. In contrast, Cromley et al. (2013) anticipated content knowledge would be necessary for answering their comprehension questions. This means that content knowledge would be less impactful on the SDCT and would not exert systematic influence across all items.
Instead, what effect content knowledge does have would be modeled as unexplained unique variance of the test items.

Finally, there is a potential statistical explanation for the discrepancy. Test questions scored dichotomously (correct/incorrect) often cluster based on distribution, or item difficulty when factor analyzed (Yang & Xia, 2015). This is particularly true when Pearson correlations are used. The use of a tetrachoric correlation matrix, unweighted least squares, and bias-corrected parallel analysis and fit statistics in the EFA minimized the risk of this occurring in the present study. Often, multiple-step items are more difficult than single-step items. Because the study by Cromley et al. (2013) was not a measure development study and the characteristics of this measure were not central to the research questions, details about their EFA methods are not reported. It is possible that Cromley et al. (2013) identified a two-factor model due to items clustering based on difficulty.

Overall, the fit of the one-factor model is reasonable. This is not surprising given the complexity of diagram comprehension as an individual difference construct. Diagram comprehension is not only affected by factors such as domain knowledge, representational familiarity, and knowledge of representations and strategies but also by the interaction of these factors (Ainsworth, 2006; Canham & Hegarty, 2010; Davenport et al., 2008; Hegarty, 2014; Kottmeyer et al., 2020a; Nitz et al., 2014; Rau, 2017). A student, for example, may be able to compensate for poor knowledge of conventions by activating an extensive prior knowledge base while another student, lacking that prior knowledge, will find it more difficult to extract information from that same diagram. In this respect, diagram comprehension ability may be similar to reading comprehension in
which strengths in some specific reading ability can compensate for weaknesses in another, such as when vocabulary knowledge supports word decoding for a reader with weak phonological skills (i.e., interactive-compensatory model of reading comprehension, Stanovich, 1984).

While additional research that can identify and disentangle these factors is necessary to better understand the processes of diagram comprehension, that is not the goal of the present study nor the purpose of the SDCT. The SDCT is intended as a test that can be used to assess diagram comprehension ability in multimedia and diagram learning research. As such, the SDCT must capture individual differences in this ability while remaining relatively brief and easily administered in the context of a broader research study. A measure that would permit the identification of the specific factors involved in the comprehension of any one diagram would be prohibitively long to address the problem the SDCT is intended to resolve. In sum, both the processes of measure development and the factor model indicate that a single total score is an appropriate way of interpreting performance on the SDCT.

**Diagram Comprehension’s Relationship to Other Variables**

Diagram comprehension is supported by several related factors and supports many academic processes. Because of the numerous diagrams in STEM education and assessment (Linenberger & Holme, 2014, 2015), the ability to comprehend diagrams is critical for learning and academic achievement (Cromley et al., 2013; Davenport et al., 2008; Shah & Freedman, 2011; Van Meter et al., 2017). As such, one would expect
diagram comprehension to be correlated with GPA. This relationship has been found in previous research (Kottmeyer et al., 2020a) and in the data reported here. The correlations between SDCT scores and GPA are low, but this is unsurprising. Several factors contribute to GPA beyond diagram comprehension, such as motivation, reading comprehension, time management, and study skills. Additionally, many classes do not rely on diagrams that contribute to one’s GPA. Overall, these results indicate that diagram comprehension is weakly related to GPA as one would expect, providing further evidence for the validity of the SDCT.

Even though comprehension is supported by a unique set of skills, most educators do not explicitly teach students how to engage with visual representations (Linenberger & Holme, 2014). Rather, instructors assume that students can comprehend and learn from visual representations (Pozzer-Ardenghi & Roth, 2005; Schönborn & Anderson, 2009). Because students are currently left to their own devices, the primary way students develop representational competence is through repeated exposure during classes as opposed to direct instruction of comprehension skills (Rau, 2017). As such, representational familiarity is a major contributor to the development of diagram comprehension (Ainsworth, 2006). This familiarity is determined by how often one encounters diagrams of a given type.

Students who are in later semesters and students majoring in the natural sciences have higher representational familiarity than beginning students or those not majoring in the sciences due to more exposure to representations. Given the importance that familiarity currently plays in developing representational competence, one would expect students more familiar with diagrams to be better at comprehending diagrams. This is
seen in the relationships between SDCT scores and both major and semester standing. While these relationships are weak, this is expected given the fact that indirect indicators of representational familiarity were used in this study. Additionally, while classroom exposure is currently the process by which individuals primarily develop representational competence, it by no means guarantees it.

Knowing that diagram comprehension can support academic processes, the next question to consider is how one can support diagram comprehension. Diagram comprehension is distinct from but related to content knowledge (Shah & Freedman, 2011), which means that one can target the ability to engage with visual representations as its own construct. This skill will be influenced by the conventions used in the diagram (Hegarty, 2014; Shah & Freedman, 2011). Representational competence is a term that encompasses the variety of specific abilities that support comprehension – even if one is unfamiliar with the content (Nitz et al., 2014; Rau, 2017). Examples include knowledge of conventions, knowledge of how diagrams function, strategy selection and use, and more (Ainsworth, 2006; Hegarty, 2014; Mayer, 2014; Sha & Freedman, 2009). As such, we would expect individuals with better representational competence to have more success in diagram comprehension (Nitz et al., 2014; Rau, 2017). The results of the present study were consistent with this expectation because knowledge about strategies, diagram uses, and ability to identify functions (RKT and RFT) were correlated with comprehension ability. Additionally, these correlations were much stronger than the relationships between SDCT scores and GPA or representational familiarity. This pattern provides even further evidence that the SDCT measures diagram comprehension ability because representational competence is more proximal to diagram comprehension than
academic achievement or indirect indicators of representational familiarity. Collectively, these relationships provide significant validity evidence that the SDCT measures diagram comprehension ability.

**Future Directions**

The SDCT exhibited a minor departure from normality, four items that consistently had low factor loadings, and a factor structure that did not explain a large amount of variance in scores. The minor departure from normality may be addressed by the addition of more difficult items to better differentiate amongst those with high comprehension abilities. However, this may not be possible due to the limitation that questions must be able to be answered solely with information in the diagram.

The low variance explained by the factor model raises some interesting questions about the nature of diagram comprehension. The combination of low variance explained and high reliability suggests that a one-factor model is a simplification of diagram comprehension. If diagram comprehension is a composite of numerous skills that interact and compensate for one another, it is unsurprising that a single factor does not explain all the observed variance. Rather, skills such as knowledge of conventions, knowledge of strategies, content knowledge, and more may explain a significant portion of variance that the one-factor model failed to explain (Ainsworth, 2006; Canham & Hegarty, 2010; Davenport et al., 2008; Hegarty, 2014; Nitz et al., 2014; Rau, 2017). As such, future work studying cognitive models of diagram comprehension should evaluate the relationship between these skills and comprehension outcomes on the SDCT. Once again, diagram
comprehension is seen as similar to reading comprehension (Kottmeyer et al., 2020a) — supported by many interacting abilities to produce a specific outcome (Stanovich, 1984).

The SDCT measures a specific outcome because including the number of questions and type of questions necessary to evaluate every possible skill that can support diagram comprehension defeats the purpose of the SDCT. Specifically, to briefly evaluate if individuals can extract information from a visual representation regardless of how. Therefore, conclusions about cognitive models of diagram comprehension are beyond the scope of the present study and the goals of the SDCT. However, future research should measure these relevant skills to study their exact impact on comprehension outcomes and possible implications for theoretical models.

**Conclusions**

The SDCT displayed good reliability, no major deviations from normality, and strong validity evidence as a measure of diagram comprehension. Additionally, scores on the SDCT relate to other variables in a way that is consistent with theory. While this provides significant empirical evidence for the validity of the SDCT as a measure of diagram comprehension, it also exemplifies the utility of the SDCT by demonstrating its relevance to academic factors and specific skills that support comprehension. The importance of diagram comprehension cannot be overstated. Visual representations play a critical role in science education (Linenberger & Holme, 2014, 2015). Diagram comprehension supports learning and problem solving (Ainsworth, 2006; Davenport et al., 2008; Hegarty, 2014; Kottmeyer et al., 2020a; Linenberger & Holme, 2014).
Therefore, it is important to support students’ diagram comprehension abilities. Despite the importance of diagram comprehension, students struggle with learning from diagrams (Cohen & Hegarty, 2007, 2012; Kottmeyer et al., 2020a). This difficulty is reflected in the low SDCT scores. Therefore, efforts should be made to improve diagram comprehension ability.

A significant barrier to supporting diagram comprehension has been the lack of a robust measure of diagram comprehension. While measures of diagram comprehension have been developed for specific studies (e.g., Cromley et al., 2013; Kottmeyer et al., 2020a; Shah & Freedman, 2011; Van Meter et al., 2017), they have had limited scope in the complexity of their diagrams, the complexity of their questions, and the domains they cover. The SDCT serves as a tool to assess students’ comprehension abilities and determine if they are improving. By evaluating comprehension as an outcome that relies on numerous skills that are applicable across numerous domains and diagram types, the SDCT measures a foundational ability: can someone extract information from a visual representation? Unfortunately, current data suggest that many students cannot. However, as a robust measure of diagram comprehension that is applicable across numerous fields, the SDCT provides an avenue to begin addressing this problem.
Appendix A

Training Materials Provided for Expert Reviewers

Please read the following instructions

In this survey, we will ask you if a series of diagrams are similar to those that undergraduate students are likely to encounter and expected to understand. Because our goal is to measure overall science diagram comprehension, similarity should not be based on the topics of the diagram such as chemistry or biology. Rather, similarity should be rated based on the difficulty and form of the diagram. An example is provided below.

There are terms we use throughout this study, please take a minute to familiarize yourself with them.

1) Difficulty: How much effort does it take to read the diagram? How many components are in the diagram?

2) Realism: Does the diagram look visually similar to what it is representing? A line graph is an example of an abstract diagram.

3) Understanding: The ability to read and extract information from a diagram irrespective of the topic it covers without support.
Characteristics of Example Diagrams

1) Difficulty: Diagrams A and C are relatively complex, with a large number of components whereas diagram B is relatively simple and easy to read due to minimal parts.

2) Realism: Diagram A and C are very realistic. Diagram A displays the physical anatomy of a reflex. Diagram C displays the various physical parts of the water cycle. Diagram B is more abstract with letters and lines holding the place of atoms and bonds.

Figure A1
Diagrams Used to Demonstrate Difficulty and Realism
Reminder:

1) **Difficulty**: How much effort does it take to read the diagram? How many components are in the diagram?

2) **Realism**: Does the diagram look visually similar to what it is representing? A line graph is an example of an abstract diagram and a blueprint is an example of a concrete diagram.

**Figure A1**

*Diagrams Used to Test Understanding of Difficulty and Realism*

Diagram A: Structure of the heart

Diagram B: A simple pendulum

Diagram C: Anatomy and actions of a beating heart
Which diagram most differs from the others in difficulty?

- Diagram A (1)
- Diagram B (2)
- Diagram C (3)

Which diagram most differs from the others in realism?

- Diagram A (1)
- Diagram B (2)
- Diagram C (3)

End of Block: Test

Start of Block: Feedback

Diagrams A and B are both relatively simple diagrams, with a small number of components. Diagram C has a large number of components, showing both structure and changes over time. Therefore, diagram C stands out as being more complex than the other two.

Diagrams A and C are both relatively realistic diagrams showing real structures while diagram B is more abstract due to its use of generic lines rather than real objects. Therefore, diagram B stands out as being more abstract than the other two.
Appendix B

The Science Diagram Comprehension Test

**Figure X:** The larva of a mosquito.

3) The abdomen of the larva __________.
   a. is the only segment with different types of hair
   b. is comprised of the eight sections labeled I - VIII
   c. contains the thorax
   d. includes the spiracular valves
Figure X: Phase diagram of CO₂

4) CO₂ will not become supercritical fluid as long as ________________.
   a. pressure stays below 10 bar
   b. temperature remains between 300 – 350 K
   c. the triple and critical points stay connected
   d. pressure stays the same as temperature

5) This diagram shows that ________________.
   a. the state of CO₂ is determined primarily by temperature
   b. CO₂ cannot transition directly from a solid to a supercritical fluid
   c. temperature and pressure combine to determine the state of CO₂
   d. CO₂ is a highly unstable gas
6) The _____________ is a period.
   a. **permian**
   b. **cenozoic**
   c. **era**
   d. **paleozoic**

7) The column labeled MYA shows _____________.
   a. how many meters deep fossils from this period are buried
   b. the number assigned to each era and period
   c. **how many years ago something happened**
   d. the difference in time between each period
8) The diagram at the top _________________.
   a. provides a close up view of the structures inside the spermatozoa
   b. is a side view of a spermatozoa
   c. includes parts of the spermatozoa not visible in the bottom diagrams
   d. provides a view of the spermatozoa from a different angle from the two bottom diagrams

9) The two diagrams at the bottom show _________________.
   a. a magnified view of the head and mid piece of the spermatozoa
   b. two different types of spermatozoa heads
   c. that spermatozoa contain either a peri-acrosomal space or a sub-acrosomal space
   d. that the centriole is located in the center of the post-acrosomal region
10) The acrosome ____________.
   a. contains the mitochondria
   b. **covers the head of the spermatozoa**
   c. is the plasma membrane that encloses the spermatozoa
   d. contains the cell membrane and nuclear envelope

**Figure X:** A desmosome is a cell structure specialized in cell-to-cell adhesion in animal cells. A type of junctional complex, desmosomes are localized spot-like adhesions randomly arranged on the lateral sides of plasma membranes.

11) Which of the following is the best label for the structure indicated by the green arrow?
   a. Keratin
   b. Desmosome
   c. **Plasma membrane**
   d. Attachment plaque

12) The ____________ are the structures that connect two cells.
   a. plasma membrane
   b. keratin
   c. extracellular space
   d. **cadherin**
Figure X: Time of fire seasons in areas of Australia

15) Which of the following is true of the fire seasons in Australia?
   a. The further south you go, the more extreme the fire season is.
   b. The fire season lasts about the same number of months in Darwin and Hobart.
   c. Autumn is when Australia is at the lowest risk of fires.
   d. Port Hedland spends more time under the threat of fire than Alice Springs.

16) This map indicates that ____________.
   a. different areas of Australia are under threat from forest fires at different times of the year
   b. any area of Australia can be impacted by fires at any time of the year
   c. the risk of fire increases during summer and autumn
   d. the number of fires differs throughout the year
24) The solid line in the figure shows ________________.
   a. the relationship between growth rates and percentage of nitrogen for grass type C
   b. that there is a difference in how nitrogen levels affect growth rates for different types of grasses
   c. how different types of grasses are affected by nitrogen
   d. that the type of grass does not have an overall effect on growth rates

25) A main idea of this figure is that ________________.
   a. grasses grow faster with higher levels of nitrogen
   b. nitrogen levels do not have an affect on grass growth rates
   c. the growth of different types of grasses is affected by the nitrogen levels in soil
   d. grass type A should not be planted in soil with high levels of nitrogen

26) Based on the figure, grass type A ____________.
   a. grows the slowest at higher levels of nitrogen
   b. grows the fastest when mixed with grass type B
   c. grows at about the same rate regardless of nitrogen levels
   d. grows the fastest at higher levels of nitrogen
Figure X: Frequency of different types of thunderstorms in the Northern and Southern regions.

27) The height differences between the blue and orange bars represent _____________.
   a. that the Northern region experiences an average of about 15 thunderstorms each year
   b. differences in the frequency of different types of thunderstorms between two regions
   c. that the number of thunderstorms varies depending on the type of storm
   d. differences in the number of single cell and supercell thunderstorms

28) The tallest orange bar shows the _____________.
   a. number of multi-cell cluster thunderstorms in the Southern region
   b. total number of multi-cell cluster thunderstorms recorded during the year
   c. difference between the rate of multi-cell cluster storms in the North and South
   d. the Southern region has about 24 thunderstorms each year
Figure X: Average temperatures for different regions of the world.

30) A main idea from the figure is that ________________.
   a. seasonal temperature patterns are different in different parts of the world
   b. temperature changes cannot be predicted by the time of year
   c. temperatures change about the same amount throughout the year in different regions of the world
   d. Ezbet Tell, Egypt and Baja California, Mexico follow the same pattern of temperature change

31) In the figure above, the labels on the x-axis indicate ____________.
   a. the major cities where temperature readings were taken
   b. the months of the year
   c. changes in the average temperature
   d. the dates when temperature readings were taken
32) What are the set of dotted lines and grey arrows intended to depict?
   a. Activation energy needed for the reaction to occur
   b. Differences in energy used in the reaction
   c. The gradual change in reaction coordinate
   d. The relationship between energy and reaction coordinate

33) Based on the diagram above, what is the difference between a reaction with an enzyme and a reaction without an enzyme?
   a. Reactions with enzymes have products with lower energy
   b. Reactions without enzymes release more energy overall
   c. Reactions with enzymes have a smaller activation energy
   d. Reactions without enzymes have reactants with higher energy
Figure X: Velocity and Position graphs, over time, for a golf ball thrown upwards

34) What is the approximate velocity of the ball when its position is 20.4 meters in the air?
   a. 20 m/s
   b. 5 m/s
   c. 2.1 m/s
   d. 0.4 m/s
Figure X: Graphs of braking and reaction distances.

35) What is the relationship between the two representations?
   a. **The graphs display the same information in two different ways**
   b. The bar graph displays the differences between the three lines on the line graph
   c. They line graph displays the total distance and the bar graph does not
   d. The reaction distances change between the two representations

36) If the speed was 32 m/s, the stopping distance would be between ______.
   a. 25-35 m
   b. 85-95 m
   c. 95-105 m
   d. **115-125 m**
37) A difference between household bleach and household ammonia is ______.
   a. ammonia is more acidic than bleach
   b. ammonia tends to be darker blue in color, while bleach tends to be a lighter blue
   c. **ammonia can have a wider range of pH values than bleach**
   d. ammonia can be used as oven cleaner, but bleach cannot

38) As the concentration of OH increases, the ______.
   a. pH of the substance decreases
   b. **substance becomes more basic**
   c. concentration of $H^+$ increases
   d. substance becomes oven cleaner
47) The thermometer is measuring the temperature of the ______.
   a. reader  
   b. stirrer  
   c. lid  
   d. water bath

48) The fuse wires are ______.
   a. connected to other wire  
   b. connected to the sample  
   c. surrounded by the bomb cell  
   d. surrounded by water
Appendix C

Representations Knowledge Test

*Note.* While a large number of questions display “A” as the correct answer, the order of the answers was randomized for participants.

1. A single visual display ____________.
   a. **can be used for more than one purpose**
   b. can only be understood in relation to written explanations
   c. is unlikely to convey important information
   d. is often all you need to understand a topic

2. When learning from a visual display, you should begin by asking:
   a. **What is this display about?**
   b. What is the reason I was given this display?
   c. There is no single best way because it depends on the visual display
   d. Is this display important?

3. The visual displays in your biology materials ____________.
   a. **are designed to serve various functions**
   b. should all be used in the same way while studying
   c. are decorative
   d. are most useful when you do not understand the text

4. Select the best resource to use if you are confused when learning from a visual display:
   a. **Class notes**
   b. Fellow classmates
   c. YouTube
   d. Textbook
5. Which is the best method for identifying the purpose of a visual display?
   a. **Identify the key elements and how they are related**
   b. Read the text elements (i.e., title or caption)
   c. Ask your course instructor for assistance
   d. There is no best way because diagrams do not serve specific purposes

6. To understand a complex visual display, you should _______.
   a. **reason about one function at a time.**
   b. read the accompanying text or class notes
   c. only focus on the most important parts
   d. skip the visual display and return to it later

7. To critically think and reason about the purpose of a visual display you should:
   a. **Ask yourself questions to guide your thinking**
   b. Test yourself on the content in the display
   c. Write a summary of what you have learned
   d. Consult the course learning objectives

8. Which of the following is not a purpose that a visual display can serve?
   a. **To show a problem-solution structure**
   b. To reason about differences in amount or magnitude
   c. To understand how the parts of a system are connected
   d. To see how two or more things are related to one another

9. In order for a display to support learning about magnitude or the amount of something, it must _________.
   a. **use some convention to display amounts**
   b. contain numbers
   c. include a legend or caption that identifies the key parts
   d. be organized in a table or graph that compares amounts

10. In order for a display to support learning about the processes of a system, it must _________.
    a. **show what is changing**
b. identify the part of the process that causes a change
c. include labels for the important structures or parts of the system
d. be shown along with some written explanation of the system

11. In order for a display to support learning about the parts of a system or structure, it must ____________.

a. show how the parts are related to one another
b. use color to show the parts of the system
c. rely on the caption or written text to explain what is being shown
d. show how the parts change in relation to one another

12. If you see a table that is included on a textbook page, then you ____________.

a. should first identify the key parts and how they are related
b. only need to skim it because it repeats what is in the written text
c. know that it is only there to show the main ideas
d. know that the purpose is to show similarities and differences between the included parts

13. Tables and graphs ____________.

a. can be used to understand how the items shown are related
b. are best used to show numbers and differences in quantity
c. cannot be used to learn about the parts of a system or structure
d. is the most effective way to learn about how a system can change over time
14. This display could be used to think about _______.
   a. the amount of solids and fluids in the body
   b. how solids and fluids change in the body
   c. how solids and fluids are related to each other in the body
   d. The redundancy of captions and legends in visual displays
15. This display could be used to think about _________.
   a. similarities and differences between types of fruit
   b. how flowers change over time
   c. the different amounts of fruit from each flower
   d. the structure of a fruiting flower
16. This display could be used to think about ______.
   a. how the structures in a plant cell are related
   b. how structures change in relation to each other
   c. the importance of color in visuals
   d. differences between plant and animal cells
17. Identify the type of reasoning you could use with this visual display:
   a. Structural
   b. Process
   c. Quantitative
   d. Organizational
18. Identify the type of reasoning you could use with this visual display:
   a. Structural
   b. Process
   c. Quantitative
   d. Organizational
19. Identify the type of reasoning you could use with this visual display:
   a. **Structural**
   b. Process
   c. Quantitative
   d. Organizational
20. Identify the type of reasoning you could use with this visual display:
   a. Structural
   b. Process
   c. Quantitative
   d. Organizational
Appendix D

Representations Functions Test

Part A: Given a task, select a visual.

**Question 1:** You are writing a lab report for your introductory biology course. In your experiment, you tested different blood samples and identified the blood type. You need to select a visual display that demonstrates the similarities and differences between the blood types. Which one of the four displays below would you select for your report?

**Correct answer:** A

---

**Blood Type**

| Red Blood Cell Types | Antibodies in Plasma | Antigens in Hemoglobin | Blood Types Compatible with Emergency
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Anti-B</td>
<td>A antigen</td>
<td>A, O</td>
</tr>
<tr>
<td>B</td>
<td>Anti-A</td>
<td>B antigen</td>
<td>B, O</td>
</tr>
<tr>
<td>AB</td>
<td>Anti-A and Anti-B</td>
<td>A and B antigens</td>
<td>A, B, AB, O</td>
</tr>
<tr>
<td>O</td>
<td>None</td>
<td>None</td>
<td>O</td>
</tr>
</tbody>
</table>

**Figure X:** Four characteristics of blood types.

**Blood Type**

<table>
<thead>
<tr>
<th>Blood Type</th>
<th>Caucasian</th>
<th>African-American</th>
<th>Hispanic</th>
<th>Asian</th>
</tr>
</thead>
<tbody>
<tr>
<td>O+</td>
<td>37%</td>
<td>47%</td>
<td>53%</td>
<td>39%</td>
</tr>
<tr>
<td>O-</td>
<td>8%</td>
<td>4%</td>
<td>4%</td>
<td>1%</td>
</tr>
<tr>
<td>A+</td>
<td>33%</td>
<td>24%</td>
<td>29%</td>
<td>27%</td>
</tr>
<tr>
<td>A-</td>
<td>7%</td>
<td>2%</td>
<td>2%</td>
<td>0.5%</td>
</tr>
<tr>
<td>B+</td>
<td>9%</td>
<td>18%</td>
<td>9%</td>
<td>25%</td>
</tr>
<tr>
<td>B-</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>0.4%</td>
</tr>
<tr>
<td>AB+</td>
<td>3%</td>
<td>4%</td>
<td>2%</td>
<td>7%</td>
</tr>
<tr>
<td>AB-</td>
<td>1%</td>
<td>0.3%</td>
<td>0.2%</td>
<td>0.1%</td>
</tr>
</tbody>
</table>

**Figure X:** Percentage of blood types in the population by race.

---

**Figure X:** Red blood cells of different blood types

**Figure X:** Graph of blood counts from lab experiment.
**Question 2:** You are writing a paper on the human heart for a project in your introductory biology course. You need to put a visual in your paper that shows the important structures of a human heart and how they fit together and are related to one another. Which one of the four displays below would you select for your paper?

**Correct answer: C**
**Question 3:** In your introductory biology course you need to prepare a PowerPoint presentation on cell membranes. You are writing a slide on the different ways molecules can pass through a cell membrane. You need to select a visual display that shows how different molecules can change location and transport across membranes. Which one of the four displays below would you select for your presentation?

**Correct answer: B**
Question 4: In your introductory biology course, you are learning about nucleotides, the building blocks of DNA and RNA. You are writing a paper on the types and frequency of nucleotides that occur in DNA. You need to select a visual that will demonstrate the frequency of nucleotides in DNA. Which one of the four displays below would you select for your report?

Correct answer: C
Functions of Diagrams Measure

**Part B: Given a visual, what might the task be?**

**Question 5:** In Sally’s biology course they are studying trees. Sally selected figure A for a report she is writing on trees. Why might Sally have selected this display?

a. To show the parts of a tree and how they are related  

b. **To show movement of water and how it is occurring**  

c. To show how much water can travel through trees  

d. To make comparisons between the roots, trunk, and leaves of a tree

![Figure A: Hydrogen bonding forms an unbroken chain of water molecules.](image-url)
Question 6: Keith is writing a report for his introductory biology course. He includes the following visual display in his report. What might be a reason Keith would include this display?

a. To show the similarities and differences between different types of cells  
b. To show how a cell changes size under different microscopes  
c. **To compare the sizes of different types of cells**  
d. To explain how cells are related to each other

Figure A: Relative sizes of biological phenomena on the same scale.
**Question 7:** Sofia is writing a presentation on mitosis and meiosis. She selected Figure A for her presentation. Why would Sofia select the following display for her presentation?

a. **To compare the stages of mitosis and meiosis**
b. To explain the structures within the cell
c. To show how many chromosomes are duplicated in a cell
d. To show the structure of the chromosomes during

Figure A: Chromosome replication during mitosis and meiosis.
Question 8: Marcus is writing a presentation for his introductory biology course. He selected Figure A for his presentation. What might be a reason Marcus would include this display?

a. To explain the structure of compact and spongy bone  
b. To demonstrate how compact bone grows  
c. To compare compact and spongy bone  
d. To show how many bone cells are in compact bone

Figure A: Cross sectional view of the inside of bone.
Appendix E

Student Demographics Survey

Q1 How many semesters of undergraduate course work have you completed?

- None, this is my first semester
- 1 semester
- 2 semesters
- 3 semesters
- 4 semesters
- 5 semesters
- 6 semesters
- 7 semesters
- 8 semesters
- 9 semesters
- 10 semesters
- More than 10 semesters
Q2 Which of the following fields best describes your major?

○ Natural Sciences (e.g., biology, chemistry, and physics)

○ Social Sciences (e.g., psychology, criminal justice, and education)

○ Humanities (e.g., history, English literature, foreign language)

○ Arts (e.g., music performance, theatre, or graphic design)

Q3 What is your cumulative GPA? If you do not have one, please leave this blank.

---------------------------------------------------------------------

Q4 How often do you complete assigned readings for classes?

○ Never

○ Sometimes

○ About half the time

○ Most of the time

○ Always

---------------------------------------------------------------------
Q5 When you are reading for a class, how much attention do you tend to give to diagrams/drawings that are in the readings?

- None at all
- A little
- A moderate amount
- A lot
- A great deal

Q6 Which of the following best describes your racial identity? Select all that apply.

- White
- Black or African American
- American Indian or Alaska Native
- Asian
- Native Hawaiian or Pacific Islander
- Identity not listed (please specify)
Q7 Which of the following best describes your gender identity?

- Male
- Female
- Non-binary
- Identity not listed (please specify)

Q8 Do you identify as having a disability that affects your ability to learn or complete classwork?

- No
- Yes
- Prefer not to say

Q9 Are you a first-generation college student? You are considered a first-generation college student if neither of your parents have earned a 4-year degree.

- No
- Yes
- Prefer not to say
References


https://doi.org/10.1016/j.learninstruc.2006.03.001


https://doi.org/10.1177/1094428114555994


https://doi.org/10.1002/acp.1344


https://doi.org/10.1016/j.lindif.2012.05.007


