

CMU-HCII-17-109

**“Alex speaks with my voice!” Promoting  
science discourse with bidialectal virtual peers**

Samantha Finkelstein

December 15

School of Computer Science  
Carnegie Mellon University  
Pittsburgh, PA 15213

**Thesis Committee:**

Justine Cassell, Chair  
Amy Ogan  
Marti Louw  
Sandra Calvert

*Submitted in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy.*

**Keywords:** Pedagogical agents, dialect, science discourse, African American English, virtual peer, culturally aligned pedagogy, peer learning, bidialectalism, code switching

December 11, 2017  
DRAFT (under committee review)

*[dedication to be filled in]*



## Abstract

As the United States becomes increasingly diverse, classrooms become home to a wider variety of student backgrounds that may not always align with traditional educator expectations. Scholars argue these mis-alignments may contribute to the systematic under-performance of students of color. In response to this, many of these scholars have also identified that *culturally-aligned learning environments* may be able to mitigate this problem and improve the performance of marginalized learners. Regardless, there remains much to understand about what specific interventions may best serve which student outcomes, and what the mechanisms are behind the success of these interventions.

At the same time, though educational technologies have the *ability* to personalized instruction, these systems have largely also treated culture as one-size-fits-all. In this thesis, we demonstrate how the design choice of one cultural factor in an educational technology, dialect, impacts the social behavior and science performance of students who speak a non-standard dialect of English. By doing so, we aim to provide additional clarity within the complex and controversial domain of culturally-aligned pedagogy. Simultaneously, by situating this research within an educational technology, we investigate how typically unquestioned decisions made in the design of technologies, such as dialect, may have unintended impacts on student behavior.

We designed a pedagogical agent that dialogues with African American students to complete open-ended science tasks. The agent was either monodialectal (spoke exclusively Standard English) or bidialectal (spoke both Standard English and African American English). Students who worked with the bidialectal agent demonstrated fewer *resistance* behaviors, like social challenging, and higher rapport. These factors, in turn, were predictive of improved post-test science performance in both experiments.

We believe this work provides three interdisciplinary contributions:

(1) These are the first studies to our knowledge that perform a controlled experimental manipulation of learning environment dialect in a domain outside of Reading Language Arts. We find that bidialectal learning environments are consistently associated with improved science performance for African American students.

(2) We demonstrate that agent dialect has significant impacts on students' social behaviors with the agent. In turn, rapport is associated with improved science performance at post-test. This adduces evidence for the hypothesis that there are quantifiable social benefits of bidialectal instruction on language minority students.

(3) We demonstrate one way in which technologies are not immune to some of the cultural critiques that have historically been given to brick-and-mortar school systems. Though perhaps unintentionally, unquestioned design choices in the systems we deploy may be unintentionally perpetuating cultural barriers that disproportionately impact marginalized students.



## **Acknowledgments**

[acknowledgements to be filled in]





# Contents

<b>Prologue</b>	<b>1</b>
Thesis overview . . . . .	2
<b>1 Introduction</b>	<b>5</b>
1.1 Language in the classroom . . . . .	6
1.1.1 Discourse and dialect . . . . .	6
1.1.2 The bidialectal approach . . . . .	9
1.1.3 Peer learning and dialect . . . . .	11
1.2 The role of educational technology . . . . .	13
1.2.1 Research questions . . . . .	14
1.2.2 Research contributions . . . . .	18
1.2.3 Thesis organization . . . . .	20
<b>2 Theoretical background</b>	<b>25</b>
2.1 Supporting language minority students . . . . .	25
2.1.1 Science discourse . . . . .	25
2.1.2 Culture, dialect, and pedagogy . . . . .	27
2.1.3 Explanations to date . . . . .	30
2.2 Dialogic learning with (virtual) peers . . . . .	33
2.2.1 Explanations to date . . . . .	34
2.3 Virtual agents and identity . . . . .	37
2.3.1 Virtual agents and dialect . . . . .	38
<b>3 Language annotation methods</b>	<b>45</b>
3.1 Pedagogical domain . . . . .	46
3.2 Language Measures . . . . .	48
3.2.1 The language of science . . . . .	48
3.2.2 Dialect density . . . . .	52
<b>4 Study 1: Dialect as an elicitation technique</b>	<b>55</b>
4.1 Study overview . . . . .	56
4.1.1 Research questions . . . . .	59
4.2 Methods . . . . .	60
4.3 Science activity . . . . .	62

4.3.1	Agent science model . . . . .	64
4.4	Results . . . . .	66
4.4.1	Science discourse . . . . .	66
4.4.2	Science arguments . . . . .	67
4.4.3	Dialect use . . . . .	72
4.4.4	Social perception . . . . .	73
4.5	Discussion . . . . .	74
4.5.1	Open questions . . . . .	76
<b>5</b>	<b>Designing a bidialectal, dialogic agent</b>	<b>81</b>
5.1	Alex: a bidialectal virtual agent . . . . .	82
5.1.1	Experimental paradigm . . . . .	83
5.1.2	Guided science activities . . . . .	84
5.1.3	Dialogue design . . . . .	86
5.1.4	Prototypical agent dialogues . . . . .	89
5.2	Rapport . . . . .	93
5.3	Dialect attitudes . . . . .	94
<b>6</b>	<b>Study 2: Interaction and transfer</b>	<b>99</b>
6.1	Study overview . . . . .	100
6.1.1	Research questions . . . . .	102
6.1.2	Science discourse assessment . . . . .	102
6.1.3	Baseline dialect assessment . . . . .	104
6.2	Overview . . . . .	106
6.3	Results . . . . .	109
6.3.1	Post-test science performance . . . . .	109
6.3.2	Intervention science and rapport . . . . .	110
6.3.3	Dialect density . . . . .	111
6.4	Secondary analyses to guide future investigations . . . . .	115
6.4.1	Standardized reading level . . . . .	115
6.4.2	Dialect attitudes . . . . .	117
6.4.3	Dialect density . . . . .	118
6.5	Discussion . . . . .	119
6.5.1	Exploratory analysis discussion . . . . .	119
6.5.2	Open questions . . . . .	121
<b>7</b>	<b>Study 3: Longitudinal impacts</b>	<b>123</b>
7.1	Research questions . . . . .	124
7.2	Methods . . . . .	124
7.3	Student case studies . . . . .	126
7.3.1	Jurnei (bidialectal agent) . . . . .	126
7.3.2	Terran (monodialectal agent) . . . . .	128
7.3.3	“Alex, do you speak Spanish?” . . . . .	132
7.4	Results . . . . .	134

7.4.1	Science discourse . . . . .	135
7.4.2	Rapport . . . . .	135
7.4.3	Social behavior . . . . .	136
7.4.4	Dialect density . . . . .	138
7.4.5	Language awareness and attitudes . . . . .	140
7.5	Discussion . . . . .	141
<b>8</b>	<b>Research discussion</b>	<b>143</b>
8.1	Science discourse . . . . .	144
8.1.1	Student responses annotated for science discourse . . . . .	146
8.1.2	Student responses not annotated for science discourse . . . . .	148
8.1.3	Limitations of Wizard of Oz approach . . . . .	150
8.2	Rapport . . . . .	152
8.2.1	“That’s what I was thinking too!” . . . . .	152
8.2.2	“How are you this dumb?” . . . . .	153
8.3	Theoretical perspectives . . . . .	155
8.3.1	Increased practice . . . . .	155
8.3.2	Second language acquisition . . . . .	156
8.3.3	Second language participation . . . . .	157
<b>9</b>	<b>Concluding remarks</b>	<b>165</b>



# List of Figures

1.1	Alex, our virtual peer experimentation platform, talking with a 2nd grade student	15
3.1	Broad feature annotation scheme for students' science language production. . . .	50
4.1	Procedure for the distant peer study across three conditions . . . . .	62
4.2	Science task images used in dialect elicitation study. Sets A and B were counter-balanced across pre- and post-test. . . . .	64
4.3	Pre and post test assessments of students' science reasoning and science discourse by condition . . . . .	68
4.4	Instances of science arguments (left), and percentage of arguments involving science discourse (right) . . . . .	71
4.5	Students' perceptions of how smart Jamie sounded when they heard Jamie speak AAE (left) or when Jamie only spoke (SAE). Though Jamie's dialect was only negatively called out in the code-switching and AAE conditions, there was no significant difference of these comments by condition. . . . .	74
5.1	Alex, our virtual peer experimentation platform, talking with a 2nd grade student	82
5.2	The three science tasks Alex works on with students . . . . .	85
5.3	Category counts of Alex's dialogue options . . . . .	87
5.4	General scheme for dialogue between the agent and student . . . . .	88
5.5	Frames demonstrating rapport at values 1 (low), 4 (neutral), and 7 (high) . . . .	94
6.1	Pre- and post- science tasks to measure reasoning and discourse . . . . .	104
6.2	Student factors by condition . . . . .	106
6.3	Relationship between hypothesized student variables . . . . .	107
6.4	The impact of condition on intervention rapport . . . . .	111
6.5	Dialect use with Alex by condition and task context . . . . .	112
6.6	Code-switching with Alex by condition and task contexts . . . . .	114
6.7	The impact of reading level and condition on intervention rapport . . . . .	115
6.8	The impact of reading level and condition on intervention rapport . . . . .	117
6.9	The relationship between attitudes about SAE and science discourse by condition	118
7.1	There was a significant impact of session on students' science argument production which seems to be lead by students' final presentations the 2nd, 4th, and 6th sessions. There is no significant impact of session after controlling for final presentations. . . . .	136

7.2	Science reasoning and science discourse during final presentations . . . . .	137
7.3	Rapport over time by condition . . . . .	138
7.4	Percentage of student utterances containing positivity and negativity over time by condition . . . . .	139
7.5	Students' dialect use between tasks and across sessions by condition . . . . .	140

# List of Tables

3.1	Morpho-syntactic SAE and AAE variants annotated . . . . .	53
-----	---	----





# Prologue

In 2013, I was observing a third grade science classroom that had become a central point of focus in the early stages of this research. The class took place in a charter school that served primarily low-income African American students. As in many urban elementary schools, the instructor, Mrs. Greenfield,<sup>1</sup> was white. She was a gregarious woman with a firm command of the classroom, and was responsible for all of the science instruction within the school.

During one of these observations, I witnessed the beginning of new biology unit that provided students with their first formal introduction to a *cell*. With a notable burst of academic urgency, 8-year-old Tyeire raised his hand and asked, “*Wait, a cell? Like a jail cell?*”

In asking this question, Tyeire astutely identified the etymological origins of the word. In the 17th century, Robert Hooke looked through a primitive microscope and described the boxy walls he observed as similar to the rooms monks stayed in, or *cellula*. In response to Tyeire’s question, Mrs. Greenfield replied, “*Stop that. We don’t talk about jail cells in the classroom. It isn’t appropriate.*”

Schools often see themselves, and are often seen, as the arbiters of what is proper, correct, and decent [Delpit, 1995]. For this reason, schools then also possess the power to communicate to young students what is *improper*, *incorrect*, and *indecent*. In dismissing Tyeire’s question about the analogous structures of a biologic cell and a jail cell, Mrs. Greenfield lost out on a potentially powerful learning opportunity. More importantly however, she communicated that a topical question about something which may be a present reality for some students with incarcerated

<sup>1</sup>All names used in the document are pseudonyms to protect anonymity.

family members is *not welcome*.

Later that year, I observed Mrs. Greenfield again during a 2nd grade science lesson about measurement. Students were asked to pick a piece of furniture around the room and use a ruler to discover its dimensions. As students scattered around the room to call dibs on various tables, Mrs. Greenfield passed out rulers to individual students. It was a mismatched collection, with each ruler of varying color or design. Darian's assigned ruler was transparent. I noticed Darian struggling with the ruler, and realized he was holding it backwards. Because of its transparency, Darian could see the numbers reflecting through from the front, but they all appeared horizontally flipped. After a minute of trying to make sense of the backwards numbers, Darian raised his hand to ask for help.

*"Mrs. Greenfield, my ruler ain't working!"*

In response to this question, Mrs. Greenfield shook her head, and turned towards the rest of the class. *"Did you hear that?"*, she asked. *"Darian said his ruler **ain't** working. What should he have said instead?"* Several students in the class chorused *"His ruler **isn't** working"* in almost-unison. *"Good,"* Mrs. Greenfield responded.

She did not respond to Darian's question, and he did not ask for help a second time.

## Thesis overview

Science literacy "for all" has been established as a national priority [United States Congress, 1994], yet in 2015, the test scores of African American fourth graders across the country fell behind those of white fourth graders by 33 points, even more than the gap in math (24) or English literacy (26). This may be in part because science often involves both. The Next Generation Science Standards for elementary school state that "every science or engineering lesson is in part a language lesson," [National Research Council, 2012] which involves both *science discourse*, the evidence styles and reasoning structures of Western dominant science, as well as Standard American English (SAE), the grammar, phonology, and lexical choices commonly associated

with “sounding educated.”

There is well documented evidence that many students who do not grow up in middle-class households with multiple generations of achieved education are less likely to come to school already familiar with the foundations of science literacy. Many may opt instead to use everyday discourse, such as explaining causal relationships through stories. Additionally, many African American students in particular are fully or partially *bidialectal*, knowing African American English (AAE) as their first dialect. When students’ language styles are less aligned with the ones required for classroom instruction, they may be prevented from reaping the benefits of education.

In this project, we investigate the ability for a *virtual peer*, Alex, to promote students’ use of science discourse. Virtual peers are a type of pedagogical agent that are designed to be partners in learning. Alex was designed on principles of guided peer learning, and models science discourse during an interactive dialogue that involves asking questions, proposing alternate explanations, and co-constructing ideas. When well supported, these techniques have been shown to help students organize their understanding through giving explanations, learn from negotiating their arguments with someone of a similar level, and so forth. Peer learning has also been linked with social benefits, with some scholars arguing it can help students find their voice in science or see themselves as a scientist.

With this foundation in place, we also use Alex to investigate one controversial theory that has been argued to be able to support the learning of bidialectal students: *dialect integration*. Scholars have argued that integrating students’ home language style into the classroom can promote students’ active participation, domain performance, and positive affect. Certain approaches have even been widely shown to promote students’ use of SAE. Despite this, many traditional science classrooms have become particularly unwavering in their insistence that *sounding like a scientist* is a non-negotiable part of *knowing science*. Students’ home dialect are rarely included (and often banned from use) during instruction. As one well-meaning educator once told me,

*“you can’t learn anything in broken English.”*

We explore the potential role that educational technologies may be able to play in supporting the educational achievement of marginalized youth in science. Though it has been over 50 years since the historical desegregation of white and African American students, we have yet to make sufficient progress in understanding how educational environments can better support the needs of these students. As educational technologies become increasingly prevalent, these systems may provide new opportunities for us to examine the conditions under which marginalized students learn best. Doing so may be able to both help us support students directly through the deployment of the intervention, as well as reveal how to best support marginalized students in classrooms more widely.

# Chapter 1

## Introduction

Each year, the United States shifts further toward becoming a majority multilingual, multicultural society [Garcia and Cuellar, 2006, Council et al., 2001] . It is estimated that more than half of all Americans will be people of color by 2050 [United States Census Bureau, 2012], and that this is already the case within an increasing majority of public elementary schools [National Center for Education Statistics, 2015]. Despite this, standardized measurements of educational achievement continue to demonstrate the purported under-performance of Black and Hispanic ethnic groups [National Center for Education Statistics, 2015]. In 2015, the test scores of Black fourth graders across the country fell behind those of white fourth graders by 24 points in math, 26 points in reading, and 33 points in science. This pervasive performance difference has long been referred to as *the achievement gap* [Jencks and Phillips, 1998, Condron, 2009, Levitt and Fryer, 2004].

Scholars have spent decades investigating the factors that may be contributing to the achievement gap, chief among them socio-economic status (SES) and access to academic resources [Rothstein, 2004, Rumberger and Palardy, 2005]. The recognition of the potential roles played by these factors are not unwarranted; one in three young black students grows up in an impoverished household [United States Census Bureau, 2017], and such conditions can multiply the prevalence of risk variables like access to health care that are also linked to under-performance [Garbarino, 1999, Harry et al., 2002]. Despite this, SES alone does not sufficiently explain the

test score gaps above. Middle-income African American students are more likely to perform similarly to *low-income* white students, rather than the middle-income white students that would be expected [Craig, 2008]. Findings such as these have led scholars to argue that there is more to the story than socio-economic status (and its associated risk) factors alone. Here, we study science literacy, the achievement gap, and the role of educational technology in addressing the issue that has dominated education policy for decades [United States Congress, 1994].

## 1.1 Language in the classroom

What exactly science literacy *entails* has been fairly poorly defined. There are a number of interactional and linguistic behaviors that have been stated as critical components to science literacy, including students' epistemological reasoning structures [Nainby et al., 2003], the ability to demonstrate dialogic inquiry with a partner [Wertsch et al., 1993, Rosé et al., 2008, Chi et al., 1989] such as by asking (and answering) deep questions [King, 1997, Chi et al., 1994], using the *appropriate* types of discourse structure [Emdin, 2010, Cazden, 2001, Michaels, 2013], and the use of specified vocabulary [Brown et al., 2005]. The Next Generation Science Standards for elementary school state that “every science or engineering lesson is in part a language lesson... from the very start of their science education, students should be asked to engage in the communication of science” [National Research Council, 2012]. In this thesis, we focus on a subset of these science language behaviors that we refer to here as *science discourse*.

### 1.1.1 Discourse and dialect

In science discourse in the classroom, children talk together to exchange ideas, information, perspectives, attitudes, and opinions. Through the dialogue exchanged between learners, students have opportunities to elaborate on their own ideas, thus both organizing their knowledge and rehearsing previously learned information. It also provides opportunities for successful construc-

tion of *new* information, such as during inquiry-based science activities where students are asked to generate their own understanding of a specific aspect within a domain. We broadly define science discourse as *the grammar of science*, or the sorts of discursive behaviors that Lemke [1990] refers to as making someone *sound like a scientist*. In many traditional science classrooms, teachers have become particularly unwavering in their insistence that *sounding like a scientist* is a non-negotiable part of *knowing science* [Lemke, 1990, Barton and Tan, 2009, González et al., 2001].

Dialogue-based learning has been associated with potential ways to afford students with a greater sense of authorship, meaning, and more equitable opportunities to learn [Resnick et al., 2015]. The features traditionally associated with dialogic learning typically entail students having opportunities to negotiate their everyday and scientific reasonings, managing alternative viewpoints, appropriating the cultural norms and discourses of the discipline, and building positive personal dispositions and identities towards science [Mercer et al., 2009, Scott et al., 2006, Boykin and Noguera, 2011].

Yet children are socialized into certain *ways of being* in their homes and communities that impact how they ask and answer questions, explain their ideas, approach academic tasks and so forth [Heath, 1983, Emdin, 2010, Michaels, 1981, Ladson-Billings, 1998]. Despite this diversity of students' home or primary dialects, schools predominantly require all communication to exclusively follow the rules of Standard American English (SAE) [Kroch and Labov, 1972], traditionally the only dialect formally recognized within most schools. African American English (AAE), by contrast, is a dialect that shares both similarities and distinctions from Standard American English. Distinguishing features of AAE include word pronunciation, grammatical constructions, phrases, intonations, lexical items, norms for politeness, turn-taking, rhythmic and stylistic timing conventions, conversational topic changes, and narrative style, and so forth [Adger and Christian, 2007, Alim, 2006, Cazden, 2001].

The predominant approach for promoting students' use of Standard American English is a

monodialectal *immersion* approach. It is rare for educational institutions to provide special programs or teaching approaches to help students acquire the standard dialect as their second dialect. Most children are taught how to read and write in the standard as if they already know it, and in some contexts, alongside children who do already know it [Siegel, 2010, 2001]. These types of approaches have also been called eradicationist approaches or deficit approaches. Scholars have argued they may have negative impacts on students. During classroom interactions, children who use non-standard dialect may be corrected by teachers in ways that lead them to withdraw into silence [Wheeler and Swords, 2004] or begin to disidentify with school [Ogbu, 1999, Sweetland, 2006, Smitherman, 1977]. Teachers communicate their negative attitudes towards students' stigmatized vernaculars through explicit denigration (e.g., "ain't ain't a word"), as well as tacitly through the use of published curricula which require the use of Standard English for an answer to be considered correct Sweetland [2006]. Across much academic work that has studied teachers' communication about language towards students, there are many published examples of teachers referring to students' language as bad grammar, broken English, ghetto language, or worse [Siegel, 2002]. Meier [1999] observes, "these are all ways of delivering the same message: what you speak is not really a language."

There is a prevalent ideology among educators that the role of school is to help prepare students for *the real world*. In this *real world*, educators recognize that there are certain ways of being, acting, and speaking that can make students seem educated, competent, and worthy of class mobility. As in the interactions described in the prologue, school is in part, or perhaps even predominantly, an institution to acculturate students into ways of being seen as *respectable*, and to relieve students of behaviors that may be counter to this goal. As Mrs. Greenfield once confided in me after one of my classroom observations, "*If children keep speaking like they belong on the streets, they're going to end up on the streets.*"

The academic research does not back this ideology. Students who do not yet fully possess fluency in SAE may demonstrate lower performance outcomes on assessments in any domain



that requires traditional knowledge of SAE, even indirectly. For example deVilliers and Johnson [De Villiers and Johnson, 2007] found that AAE-speaking students who are not yet fully fluent in SAE may struggle to understand the linguistic rule of *third person singular -s* in ways that negatively impact sentence comprehension in some academic tasks. Similar results are reported by Terry and Smith [Terry et al., 2010], who find that when students do not demonstrate fluent comprehension of third person singular -s, they are more likely to score poorly on math problems that involve the use of this linguistic feature. The authors argue for a cognitive load explanation for this result, indicating that the linguistic mismatch added additional challenges necessary for completing the task. Additionally, classroom-facing ethnographic investigations have served a critical role in helping to make visible differences in student experiences that might otherwise go unnoticed [Rogoff, 1990, Michaels, 1981, Gutiérrez-Clellen et al., 2004]. These and many other studies will be detailed in subsequent chapters of this thesis.

### 1.1.2 The bidialectal approach

With these results in mind, the teaching of science discourse runs headlong into classroom beliefs around dialect and *cultural synchrony*, the complex and integrated factors that have been identified to contribute to students' classroom interactions. Many scholars have long-argued for learning experiences that integrate in aspects of students' home culture, including their language system [Fradd and Lee, 1995, Brown and Spang, 2008, Lee, 2003, Boykin, 1994, Seiler, 2001, Ladson-Billings, 1995a, Au et al., 1981]. These interventions have been global and varied, but largely share several similarities. Broadly, they are focused on supporting the achievement of historically marginalized populations, and they do so by *lessening the rigid barriers* that so frequently separate students' home culture from the cultural expectations of the classroom.

*Code switching* or *dialect shifting* is the process of shifting between styles of language based on *context*, which could be impacted by speech partner, topic of conversation, conversation setting, what groups the speaker is looking to align (or dis-align) with, and so forth. There is some

evidence that aligning your own dialect with someone else's within the same linguistic community can serve positive social functions, and that girls are more likely than boys to demonstrate this type of accommodation [Van Hofwegen and Wolfram, 2010, Van Hofwegen, 2015]. Studies of intra-person dialect use have consistently found that use of AAE features are not random, and likely serve communicative functions like signaling group identity, denoting emphasis, adjusting the rhythm and *feel* of a phrase, or aligning with a particular idea.

Research by Charity et al. [2004b] demonstrated that for 217 African American children from kindergarten to second grade, greater familiarity with SAE corresponded to higher achievement in reading as measured by standardized tests. Craig and Washington [Craig and Washington, 2004] found that students who reduced their AAE in school and shifted to SAE performed six times better on standardized reading tests than those who maintained AAE use and did not shift. This shifting towards SAE was also associated with positive results on tests of vocabulary development.

Analyses from Terry and Scarborough [2011] identified that students often knew the standardized English form for words, even when they were not employing these forms. For example, in two studies with children in preschool - second grade, these authors found that children who spoke AAE frequently displayed considerable knowledge of SAE forms on non-word repetition, picture naming, and naming judgment tasks. That is, children who used AAE in overt speech had precise lexical knowledge of phonological SAE forms. This knowledge, regardless of its use or lack thereof in a given exchange, was found to be positively associated with academic achievement.

The research above reveals an important distinction: in traditional school systems, it is not that the use of AAE is a *negative* predictor of success; rather, it is that the use of or knowledge of SAE is a *positive* predictor of success. In other words, if students are able to demonstrate knowledge of and use of SAE in the contexts in which it is required, their use of AAE in other contexts does not negatively predict performance. In fact, the linguistic flexibility to use both

dialects based on context seems to be associated with greater academic performance than among African American children who spoke consistently low levels of AAE [Craig et al., 2009b]. For this reason, many scholars have focused their efforts on *promoting students' use of SAE* without *eradicating their use of AAE*. In other words, many of these approaches have focused on promoting students' demonstrations of code switching, and have shown themselves effective. In the next section, we explore the implication of these results on discourse, peer learning, and their impact on the achievement gap in science literacy.

### 1.1.3 Peer learning and dialect

Often, peers are perceived to have either high or low *social capital*, an invisible imagined amount of social currency (such as popularity, or lack thereof). *Social impact theory* suggests that popular peers are especially influential and may have a disproportionate impact on how likely their own behaviors are to be emulated by others [Cillessen and Rose, 2005]. Conversely, *social misfit theory* posits that students who stray from the norms of social expectations, especially in a way that marks someone as low status, may promote both social or physical aggression [Dodge et al., 1990, Farmer and Rodkin, 1996]. This theory posits that while demonstrating the behaviors that are a mark of in-group behaviors may not consistently reward students with social capital and group acceptance, demonstrating *low-status behaviors* may result in students responding with *rejection* or *repulsion* [Wright et al., 1986].

From this angle, the dialogue *partner* in a learning exchange may differently impact the students' *own* behaviors based on factors like the partner's perceived social status [Latane, 1981], similarity to themselves [Byrne, 1971], or the level to which they expect that emulation of the partner's behaviors will grant them social capital [Cillessen and Rose, 2005]. According to these perspectives, we may be *more likely* or *less likely* to engage in an interaction with a partner based on who a given student is, who they perceive the partner to be, and what they believe they might have to gain or lose, socially, from interacting with this particular in a particular type of way.

This moderates the impacts of how effective a peer learning exchange might be.

When we deem someone *like ourselves* in a salient way, we may be more likely to feel a sense of belonging [Meyerhoff, 1996]), choose them as friends [Byrne, 1971], accommodate our speech patterns to be better aligned with their own language use [Abrams et al., 2002, Fought, 1999, Eckert, 2008], or emulate their behaviors [Berndt, 1999, Berndt and Keefe, 1996, Jones et al., 2012]. In learning contexts, these broad concepts of positive social environments have been associated with being a supportive environment for learning [Frisby and Martin, 2010, Madaio et al., 2016, Palinscar and Brown, 1984].

From this angle, the dialogue *partner* in a learning exchange may differently impact the students' *own* behaviors based on factors like the partner's perceived social status [Latane, 1981], similarity to themselves [Byrne, 1971], or the level to which they expect that emulation of the partner's behaviors will grant them social capital [Cillessen and Rose, 2005]. According to these perspectives, we may be *more likely* or *less likely* to engage in an interaction with a partner based on who a given student is, who they perceive the partner to be, and what they believe they might have to gain or lose, socially, from interacting with this particular in a particular type of way. The following theories have broadly been argued to moderate the impacts of how effective a peer learning exchange might be.

Assessments of social influence moderators like popularity, attractiveness, romantic interests, or interpersonal peer group clique tensions have rarely been considered in traditional peer learning investigations. And yet, for anyone who has been in a grade school classroom, it seems intuitive that students' own social agendas are at the very least *just as likely* to influence their behaviors as the goals their educators hold for them [Brown et al., 2017a]. Though much remains to be understood about how these processes work and why exactly they might moderate students' behaviors, evidence exists that there is some impact on how students carry themselves in peer learning exchanges. For example, rapport and friendship status have been evidenced to be correlated with how dialogic behaviors like feedback, hints, and instructions are provided dur-

ing the exchange Hartup [1996], Parr and Townsend [2002], Madaio et al. [2017]. Conversely, when students feel socially threatened by a learning partner, they may develop “non adaptive” behaviors that limit the potential for learning [Newman, 2002].

In light of all this, we must hedge the praise given to peer learning opportunities as being simply that: *opportunities*. Largely, studies of the success of peer learning opportunities has found that students *do not know* or *do not choose to* engage with an other in the sorts of ways that are associated with academic success. Students working in groups may be more focused on just identifying the right answer rather than demonstrating the sorts of discourse patterns which are associated with achievement [Vedder, 1985]. Similarly, though the above section lauded the potential for these environments to promote students’ thoughtful elaboration, such idea justifications are typically only elicited by thoughtful *questions*, and students don’t simply come with the ability to collaborate well *off the shelf*.

## 1.2 The role of educational technology

The driving motivation of this thesis is the investigation of how theories of dialect-integrated instruction may also have pedagogical implications in the design of educational technologies. Because the experimental variables we investigate in this work (science discourse and dialect use) are both based on language, we situate our intervention in a dialogic peer learning design.

A pedagogical agent is any character with a visual presence in the learning environment that is specifically designed to facilitate learning [Johnson et al., 2000, Moreno and Flowerday, 2006, Schroeder et al., 2013, Schroeder and Adesope, 2014]. There are many ways agents can aid the learning process, from specific actions such as signaling [Craig et al., 2015, Johnson et al., 2013b], through motivational speech [van der Meij et al., 2015], playing different roles as a facilitator in the environment [Clarebout et al., 2002] or modeling learning strategies [Craig et al., 2012, Twyford and Craig, 2017]. While initial systematic review indicated no differences in learning with agents [Heidig and Clarebout, 2011], a more recent meta-analysis has shown

that pedagogical agents can facilitate a significant, small positive effect on learning outcomes [Schroeder et al., 2013]. However, researchers have also shown that the design of the pedagogical agent, meaning its voice, speech patterns, or outward appearance, can influence how the agent is perceived or how effectively the agent facilitates student learning [Baylor and Kim, 2004, 2009, Clark and Choi, 2005, Domagk et al., 2010, Kim and Wei, 2011, Moreno and Flowerday, 2006, Ozogul et al., 2013, Schroeder et al., 2017, Veletsianos, 2010]. These results highlight the importance of purposeful, data-driven agent design.

We aimed to design an educational intervention that we expected to support student learning based on principles of guided peer collaboration, as well as allow us to easily manipulate the system’s dialect within an appropriate social context. Vygotsky’s widely cited initial conceptions of learning from a *more advanced other* [Vygotsky, 1982] has inspired technologists to consider how technologies might be able to play the role of that more advanced other.

We hypothesized that if agent dialect had no impact on student performance, as is the default hypothesis in most technological learning environments that provide exclusively monodialectal learning, students would learn from our pedagogical agent regardless of its dialect. Conversely, if our analyses revealed that agent dialect plays a moderating role on student performance, it would lend evidence to the argument that unquestioned culturally-marked design decisions in educational technologies, such as language, may have unrealized impacts on the success of marginalized students.

### **1.2.1 Research questions**

We answered the questions presented in this thesis through the use of a pedagogical agent named Alex. Alex operates through a *Wizard of Oz* approach where a hidden experimenter listens in and chooses Alex’s response out of over 1,000 pre-recorded options. We do this to avoid barriers introduced by state-of-the-art automated approaches, which are limited both in social fluency and speech recognition for young students who speak non-standardized English dialects.



Figure 1.1: Alex, our virtual peer experimentation platform, talking with a 2nd grade student

We ran two investigations. In both studies, each session with Alex lasts about 20 minutes. In the first half, the pair brainstorms ideas in an informal, open-ended peer dialogue. In the second half, the pair takes turns pretending to be a teacher, asking and answering review questions about the material. In each study, we manipulated the dialect pattern Alex demonstrated. In one condition, Alex was *monodialectal* (that is, only spoke Standard American English). In the second, Alex was *bidialectal*, speaking African American English during the peer brainstorming task and then *code switching* to Standard American English for the teacher presentation task. In other words, during the peer brainstorming task Alex spoke using SAE in the monodialectal condition and AAE in the bidialectal condition. During the teacher presentation task, Alex spoke using SAE in both conditions.

Because of the growing body of research indicating that guided peer dialogues with a virtual agent can promote students' learning, we hypothesized that Alex would successfully be able

to promote students’ use of science discourse. However, because previous research with pedagogical agents has found that agent design plays a large role in how effective agents are, we expected that Alex’s dialect might result in students’ demonstrating different behaviors towards Alex during an intervention, and perhaps because of this, different levels of science discourse at post-test.

There are many questions that could be answered with this data. For focus, I scoped our primary analyses to the questions below.

1. After interaction with a pedagogical agent that models science discourse, do students show improvement in their own science discourse at post-test?
2. Do students who work with a bidialectal agent demonstrate *greater* improvement in their use of science discourse at post-test than those with a monodialectal agent?

These first two questions serve to make two contributions to the pedagogical agent community in ways that may be applicable to educational technology more broadly. The first question investigates the impacts of a pedagogical agent learning environment on students’ improvements to their science discourse. The second explores whether manipulating an often-unconsidered design choice, *dialect*, may impact the effectiveness of this system for language minority African American students.

There has been much discussion within educational technology communities, including scholars who work with pedagogical agents, about how to support under-served students using these systems. Often, the benefits of these systems are argued to be in their ability to increase access to personal tutoring for those who cannot afford it, or in their ability to promote the sorts of cognitive benefits hypothesized to be associated with learning in a “culturally neutral” way (e.g., *immediate feedback is useful*). Despite this, additional evidence in the agent community is demonstrating the ways in which design impacts student behaviors. In answering this second question, this work contributes to our understanding about how intentionally designing for cultural variations that are supported in the literature might translate to technology and promote the



benefits for students using the system.

Second, we investigate the hypothesized theory that one of the benefits of bidialectal learning environments are the benefits to socio-emotional well-being. Scholars have argued that culturally aligned learning environments work in part through their ability to promote a sense of community, or improve students' identification with the classroom learning environment [Brown et al., 2017b]. Similar claims have been made for the benefits of peer learning more broadly. In both cases, there remains much to understand about the potential role positive social behaviors play in students' learning experiences. This leads me to my next two questions:

3. Do students with the bidialectal agent demonstrate more rapport during the intervention?
4. Does increased rapport during the intervention lead to greater use of science discourse at post-test?

To our knowledge, there have not been manipulations of dialect in a learning environment that was not focused on helping students learn Standard American English. Specifically, this is the first series of studies to investigate how a bidialectal pedagogy can promote students performance in science achievement and the language of science literacy. In doing so, we may be able to demonstrate another way to respond to the call from Lemke [1990] to "let kids talk in their own language." This call originally aimed to invite everyday reasoning into the science classroom; here, we explore the separate impact of students' home dialect in their classroom science interactions.

We focus our investigations primarily on differences that emerge during a presentation task. This allows us to separate the peer task, where students interact with Alex in ways that vary by condition, from the more controlled teacher task, where Alex's dialect was identical in both conditions. Thus, we expected any difference in students' performance during the teacher task to be due to the dialect Alex demonstrated in the previous task.

Even if this difference is shown, social reactions to a technology intervention can have simpler explanations. For instance, the *novelty effect* of interacting with technology can produce

short-term gains but vanish over time [Kiesler and Sproull, 1997]. Bidialectal pedagogy may also accelerate familiarity with a new agent, presenting new concepts, but that effect may vanish over time and converge with results from a monodialectal approach. Our final research question aims to eliminate these possibilities through the exploration of bidialectal pedagogy over time.

5. Can novelty or familiarity effects be excluded as insufficient explanations for the impact of bidialectal pedagogy?

Finally, we identified hypothesized moderating variables due to the social theories overviewed in later chapters, such as the identification that low-ability students respond differently than high-ability students to agents who seem particularly enthusiastic about doing the task, or that perception of in-group status might impact how likely students are to accommodate or emulate partner behaviors. This opens up an extended research agenda for future research on educational technology and the impact of bidialectal pedagogy in the classroom. Though we cannot make strong claims about the results identified in this area due to a limited sample size and the limitations of running multiple analyses, we present the results of several guiding findings that could guide future scholars to consider these student factors more directly in their own work.

### **1.2.2 Research contributions**

The work presented in this thesis demonstrates that the dialect used by a pedagogical agent has implications for the social behavior, and in turn the science performance, of language minority African American students. We find that this result maintains over a six week longitudinal study. We find that while students generally demonstrate within-subjects improvement regardless of Alex's dialect, these gains are greater for students with the bidialectal agent. This finding seems to stem from our results in the two virtual peer experiments which demonstrate that students with the monodialectal agent are more likely to demonstrate outwardly negative social challenging behaviors such as through insults, aggressions, and threats. We interpret these findings throughout this thesis from both cognitive and sociocultural perspectives to add clarity about the relationship

between the dialect used by a pedagogical agent, students' social responses, and ultimately their achievement within the task. In interpreting these results, we consider one additional finding that emerged in this work: agent dialect also seemed to impact students' self-reported language ideologies during a post-test assessment. Students with the monodialectal agent were more likely to demonstrate more negative attitudes about Standard American English at post test.

I believe this work provides three interdisciplinary contributions:

(1) These are the first studies to our knowledge that perform a controlled manipulation of the dialect used within an educational technology to support the achievement of language minority African American children. To my knowledge, it is also the first set of experiments that have investigated the benefits of a bidialectal pedagogy approach towards a domain outside of Reading Language Arts. We find that bidialectal learning environments are consistently associated with improved science performance for African American students.

(2) We demonstrate that agent dialect has significant impacts on students' social behaviors with the agent. In turn, rapport is associated with improved science performance at post-test. This adduces evidence for the hypothesis that there are quantifiable social benefits of bidialectal instruction on language minority students.

(3) We demonstrate one way in which technologies are not immune to some of the cultural critiques that have historically been given to brick-and-mortar school systems. Though perhaps unintentionally, unquestioned design choices in the systems we deploy may be unintentionally perpetuating cultural barriers that disproportionately impact marginalized students.

This work makes contribution to two primary communities. To the pedagogical agent literature, we find that manipulating agent dialect can promote significant improvements to students' performance in science literacy. Culturally aligned technology can be conceptualized in more deep-seated ways that support students in their growth as learners. To the broader education community, this work provides additional clarity on the success of bidialectal pedagogy in supporting underserved student groups. In addition, this work demonstrates one way in which peda-

gological agent interventions can be used to evaluate hypotheses that have been theorized in social science literature, but that are hard to evaluate in more controlled settings. This work serves as a call to educational technology communities to consider the design choices being included in their systems, and whose culture is being privileged through its choice; at the same time, it is also a call for technologists to be *using* these systems to investigate hypothesized variables that have been intended to support the improved performance of marginalized youth.

### 1.2.3 Thesis organization

In this thesis, I present three investigations that aim to provide clarity to the debates surrounding the ways unquestioned design decisions about language in an educational technology might moderate the learning experience for language minority students. Scholars argue that culturally sustaining pedagogies, particularly ones that integrate students' home dialect, may be able to support the academic performance of language minority students. Despite this, learning environments largely demonstrate the exclusive use of dominant language practices. In this work, we examine how students' learning experiences may be impacted by the dialect used in the design of one type of educational intervention called a virtual peer. If learning environment dialect is irrelevant for students' learning, we would expect students to learn equally with the agent regardless of its dialect; conversely, if there are observed differences that seem to be associated with agent dialect, examinations of this data may reveal what mechanisms may be driving this observed differences, and what student factors may moderate the extent to which this choice holds weight.

Throughout the remainder of this thesis, I present the background research from prior scholars that has motivated our investigation of language as a potentially important factor in students' learning [Craig et al., 2009a, Wertsch et al., 1993, Reveles et al., 2004, ?], the use of virtual agents as experimental platforms [Moreno and Flowerday, 2006, Veletsianos et al., 2008], and the pedagogical principles on which we based the design of our virtual agent as a way to promote its efficacy as an educational intervention [King, 1994, 2002, Rogoff, 1990, Lee et al.,

2013, 2005]. I describe three studies which were completed in order to answer the question of whether or not agent dialect impacts the efficacy of a learning environment. This work bridges multiple “fields” and disciplinary boundaries, and I aim to review them each below so as to be thorough in my explanations of both the motivations and interpretations of this work. For clarity, the outline below describes the order in which I present the information in this thesis, as well as the argument flow this thesis intends to follow.

**Chapter one: Introduction.** There remains a difference in standardized achievement between White and African American students. Scholars have argued that culturally sustaining instructional techniques may hold benefits for marginalized students. Despite this, the majority of learning environments (both physical and digital) are designed to be “culturally neutral.” In this work, we investigate how manipulating the design of one often-unconsidered but culturally-marked design choice, dialect, might impact the effectiveness of an educational technology for language minority African American students. We answer this question using a virtual dialogic agent designed to promote science discourse. Virtual agents have both been shown to be able to support student learning as an intervention, as well as serve as an experimentation platform for controlled manipulations of social factors. For these reasons, dialogic virtual agents may be a compelling platform to investigate the potential impacts of technology dialect on student performance.

**Chapter two: Background.** In this chapter, I review the social and cognitive mechanisms that have been associated with the benefits of peer learning approaches, pedagogical agents, and language aligned learning environments. I draw on these concepts towards my interpretation of the presented results.

**Chapter three: Measures.** I define here the two language variables which we most directly target in this work: science discourse and dialect use.

**Chapter four: Study 1.** Before building our agent, we performed a simple “manipulation check” study with a non-interactive agent. This allowed us to evaluate our conceptions of dialect and science discourse, and collect student language data that would allow us to build the design of our subsequent dialogic agent. This study revealed that students who had heard a non-interactive “worked example” of science discourse in African American English demonstrated approximately twice the amount of science discourse as those who heard the same content from an agent who exclusively spoke using Standard English. This study inspires our next questions: what driving mechanisms, either based in cognitive or sociocultural factors, may best be able to explain why this might be?

**Chapter five: Methods.** In this chapter, we introduce the research questions that motivate our following two studies: what can analyses of the dialogues between the student and an agent reveal about what impacts technology dialect has on student performance? I here present the design of our virtual agent Alex, and how we chose an intervention paradigm which we hoped would be able to reveal insights about this relationship.

**Chapter six: Study 2.** In this study, 30 African American 2nd graders interacted with either a monodialectal or bidialectal version of Alex for one 20 minute session, and were given post-test science assessments the next day. This study revealed that students with the bidialectal agent demonstrated more rapport with Alex during the intervention, as well as greater improvements to their science discourse on a next-day post-test.

**Chapter seven: Study 3.** In this study, we present a longitudinal six-week investigation with 11 students to understand how agent dialect might have different impacts on students over time. This study was done with the recognition that after six weeks, a potential *novelty effect* may

wear off, resulting in the lessened performance of those with the bidialectal agent, and thus no difference between conditions. Conversely, we posited that we might also find that students with the monodialectal agent might eventually *catch up* to those with the bidialectal agent, such as if it simply took them longer to build rapport. However, our analyses revealed that there were differences in social behavior by condition throughout the six week experiment. We found salient differences in the amount of “social challenges” (sometimes referred to as “agent abuse”) by condition. Students with the monodialectal agent demonstrated almost ten times the amount of these resistance behaviors on average across the six week intervention. These behaviors were most commonly demonstrated by students who reported less positive attitudes about Standard English at pre-test.

**Chapter eight: Discussion.** In this discussion, I look at the three studies presented in this thesis holistically and identify what mechanisms may be best able to explain these results. Because there were differences in the exact implementation of the dialogic agent, I aim here to be careful not to overstate any potential findings. However, across both studies with the agent, the primary take-away of our results appears to be that agent dialect does impact the interactions of some students during an intervention, and that these behaviors are in turn associated with post-test performance. Brown [Brown et al., 2017a] argues that science discourse may carry *ideological baggage* for some under-performing African American students, and Fordham [Fordham, 1998] argues that students who are faced with socially threatening academic learning environments may respond with resistance behaviors that distance themselves from participation. Our findings adduce evidence for the possibility that perhaps a bidialectal agent mitigates this face-threat for under-performing students. It might also be the case that the monodialectal agent actively contributes to this face-threat by representing a character who might “look like the students,” but not demonstrate behaviors that might for some students not feel culturally appropriate.

**Chapter nine: Concluding remarks.** I conclude this thesis by opening questions to education researchers about the role that culture plays in our designs of learning environments. What are the steps that must be taken for learning scientists and designers of educational environments to assure that their systems are promoting educational equity? I conclude that the choice of language in an educational technology, more often simply *having not made a choice at all*, may impact the social reception of the intervention and, thus, its effectiveness. For these reasons, I argue that factors may be often dismissed in analyses of data, such as disengagement or agent abuse, may not be *noise*, but rather an indication that there are students' whose social needs are not being supported. By recognizing those students, educational technologies can support a global mission towards equity.



## Chapter 2

# Theoretical background

As this thesis engages with the interwoven effects of dialect, science discourse, and educational technology in the classroom it is useful to separate the issues and describe our current understanding of each topic. This chapter surveys each topic and provides a crosswalk to identify bridging research between subfields.

## 2.1 Supporting language minority students

### 2.1.1 Science discourse

Lemke, a prominent scholar of science education, argues that our narrow techniques for promoting science among students have been historically ineffective [Lemke, 1990]. He argues that part of our difficulty in helping to teach students science discourse is in our recognize the ways in which it is a type of *second language*. He argues that the language of science may feel foreign and uncomfortable to most students, and that allowing a bridge between students' colloquial language and the language of science may be able to promote achievement.

Following Lemke's call, some scholars are beginning to examine the impact of helping students make connections between their own colloquial language and science language [Brown,

2004, 2006, Reveles et al., 2004, Varelas et al., 2002]. For example, Brown et al. [2017b] investigated computer-mediated science learning experiences that either exclusively used the language of science, or first taught students science concepts and *then* introduced the rules of science discourse. These investigations on elementary school African American students found that those who were only asked to use science discourse *later on* demonstrated better science domain knowledge and an increased use of science discourse.

Brown refers to these approaches as having *disaggregate instruction*. Disaggregate approaches to science learning are successful for two reasons. First, by separating the conceptual and language components of science, students may be better able to understand the concepts behind the terms being used, and then learn how to translate them into science discourse. Second, Brown et al. [2017b] argues these approaches may be able to support students in shaping their attitudes and identities within science.

Similarly successful approaches were also identified by Reveles et al. [2004], who examined a teachers strategic approach to showing students a vision of how their everyday terms were similar to those used in science. The teacher in their study promoted students use of science discourse by drawing parallels between their everyday language and the language of science. Students were given opportunities to explain and discuss science ideas in everyday terms, and were then asked to *translate* them into science discourse. The works above, as well as Emdin [2016], demonstrate that African American students benefit from learning environments that integrate students' more comfortable language styles into science pedagogy.

Science discourse interfaces with student perceptions of identity and dialect. Brown et al. [2017b] introduces one interpretation for why science learning interventions might be successful: "people view language as symbolic of cultural membership; therefore, when they encounter alternative languages, these encounters may signal identity mismatch". This matches Lemke [1990]'s argument that science discourse is embedded within a larger set of cultural practices and values; for this reason, speaking science discourse may involve the *taking on* of that *cultural*

*baggage.*

### **2.1.2 Culture, dialect, and pedagogy**

Students who do not speak SAE at home come to school as second dialect learners, with separate but interlocking ways in which they are disadvantaged in that new environment [Siegel, 2010]. First, they have to read and write in a language variety that they do not already know (in this case, SAE), and they have to learn content in subjects such as mathematics in that variety as well. Second, since they are not considered to be language learners, they are not given any special instruction to help them achieve fluency in the target dialect. Third, they may become frustrated by this type of correction and their inabilities to express themselves comfortably during any classroom interaction.

These disadvantages are at odds with monodialectal classroom teaching, the prevalent pedagogy today. The primary argument given by educators in response has been referred to the *time on task* hypothesis [Cummins, 2000], with the argument that the greater amount of time instruction is in SAE, the better students will demonstrate use of that dialect. The alternate, more firmly stated argument for monodialectal pedagogies is that use of students' home dialects *interferes* with students' abilities to fully learn the target dialect. In other words, the *interference* hypothesis provided by educators is that it would be too confusing for students if they are allowed to use both language systems, and that monodialectal learning environments can help students better recognize the boundaries of SAE [Siegel, 2010].

Student beliefs vary. Some, but not all, African American students view AAE as an important and valuable language style [Godley and Escher, 2012, Ogbu, 1999, Sweetland, 2008]. For example, Godley and Escher [2012] identifies that high school African American students who are demonstrating high levels of academic success within the school system may be less likely to both use as well as hold positive attitudes about AAE.

In contrast to primarily “monocultural” learning experiences that are common in traditional

classrooms, bidialectal approaches to achievement have most commonly taken the form of integrating in aspects of students' home dialect into instruction towards the target dialect (in this case, SAE), and they have primarily been investigated in English classes.

These approaches have been shown to be successful for promoting school-ratified literacy among a number of language minority groups. Au et al. [1981], for example, conducted a case study of the Kamehameha Early Education Program (KEEP), a language arts development project, to examine how to best teach native Hawaiian children to read. Au and Jordan found KEEPs method of reading instruction to be successful in part because of how reading lessons were conducted. The instruction was designed to incorporate the linguistic patterns frequently found in native Hawaiian culture, namely *talk stories*, which focused on co-construction and social relevance.

Similarly, Mohatt and Erickson [1981] similarly found that Native American students demonstrated improved academic performance when teacher-student interactions incorporated “mixed forms” of language. When teachers demonstrated a combination of Native American and Anglo language patterns, students demonstrated increased participation and increased learning outcomes.

Ladson-Billings [1995b] uses the phrase “culturally relevant teaching” in her descriptions of what makes successful teachers of African-American students. She describes this approach to teaching as one “that empowers students intellectually, socially, emotionally, and politically by using cultural referents to impart knowledge, skills, and attitudes.” Through this framing, she points out that bidialectal approaches to literacy instruction may be beneficial because they support students in both the cognitive *and* social factors that may be associated with students' use of SAE.

Some authors have suggested empowering students to use their linguistic expertise through inquiry-based instruction [Wolfram and Christian, 1989]. Alim [2009] led students through Critical Hip Hop pedagogies, in which they analyzed the music style for linguistic variation.

Chisholm and Godley [2009] taught Critical Language Awareness strategies through student inquiry-based instruction. In this work, students used their metalinguistic awareness to discuss language variation, identity, and power in the United States. These instructional techniques have been associated both with improved student performance as well as improved student motivation.

### **Contrastive analysis**

Many scholars have specifically focused on bidialectal pedagogies not just to promote student achievement and engagement more generally, but also to promote students' use of SAE. These approaches have largely involved a technique called *contrastive analysis*, which involves helping students translate between SAE and AAE, and have largely demonstrated success [Wheeler and Swords, 2004, Fogel and Ehri, 2000, Sweetland, 2006]. For example, Taylor [1989] designed a writing curriculum designed to teach SAE grammatical structures. In the *business as usual* classrooms, the composition curriculum taught ten Standard English features explicitly. In the experimental condition, he used an approach called *contrastive analysis* that compared and contrasted features of SAE with AAE. At the conclusion of the eleven week curriculum, students in the experimental group demonstrated a reduction of AAE features within their composition by 58%. Conversely, those in the *business as usual* condition showed a significant *increase* in their AAE use by 8.5%.

Sweetland [2008] performed a similar experiment with classrooms of African American elementary school students that involved a contrastive analysis approach. The writing curriculum also included class time devoted to reading books and listening to audio recordings of other non-standardized language varieties, such as Southern English. Classroom teachers were randomly assigned to either teach using this culturally relevant curriculum or a traditional one that removed the *marked* references to language diversity.

Sweetland demonstrated that students in a culturally aligned curriculum demonstrated much stronger use of SAE on their writing assignments, as well as generally writing longer essays. She

also identified that students seemed warmer and more engaged throughout the ten week period over which her curriculum was deployed. An important results from Sweetland's study was that through the measuring of teacher attitudes in both conditions before and after implementation of the curriculum, teachers who had taught using the culturally aligned curriculum also reported more positive attitudes about AAE *themselves*.

Fogel and Ehri [2000] designed a large-scale study involving twelve 3rd and 4th grade classrooms of African American students that also investigated the contrastive analysis approach. Classrooms were randomly assigned to receive one of three types of writing instruction: (1) exposure to SAE grammatical features, (2) exposure to SAE and explicit instruction on how to identify features of SAE, or (3) both of the above interventions, as well as guided practice translating AAE grammatical structures into SAE. Students who additionally received contrastive analysis guided practice demonstrated significantly more instances of SAE features and a better ability to translate sentences into SAE at post-test. These authors emphasize that there was no difference in students' use of AAE features from pre-test to post-test when students exclusively received exposure to Standard English. The authors argue that teaching linguistically-diverse students SAE requires guided practice and immediate feedback, and that contrastive analysis techniques appear to be particularly successful at accomplishing this goal.

### **2.1.3 Explanations to date**

Though there's evidence that bidialectal pedagogies may support student achievement, much remains to be understood about *why this might be*. The benefits of bidialectal pedagogy have been attributed to improving students' comprehension of instructional material [Nieto, 1999, Cummins, 1979], improving their metalinguistic awareness of their own dialect production [Edwards et al., 2010].

As classroom-level experiments demonstrate, the mechanisms through which bidialectal pedagogies support students' use of SAE might either be cognitive (such as helping them better iden-

tify linguistic differences through contrastive analysis) or socio-emotional (such as improving students' perceptions of the classroom climate, making them feel culturally validated, improving their desire to participate, and so forth) [Lee et al., 2005, Sweetland, 2008, Wheeler and Swords, 2004, Ladson-Billings, 1995b, Boykin, 1994, Siegel, 2010]. As Ladson-Billing's argues herself in her initial conceptions behind the success of culturally relevant pedagogies, these curricula *should* be designed to support students through both cognitive and sociocultural mechanisms. However, for these reasons, bidialectal or culturally aligned pedagogies may seem in part like a *black box*. From a perspective of praxis, these interventions have demonstrated the potential for bidialectal approaches like contrastive analysis in supporting student achievement. From a research perspective, additional work to better understand what factors might mediate the relationship between these interventions and student performance may help us better design the most promising interventions for students who are not currently achieving in traditional learning environments. In the same ways, it is also important that the work of scholars designing culturally relevant curricula do not take *essentialist* perspectives that consider all students from language minority groups equally at-risk of under-performance, nor that all students may benefit equally from the same approaches (or even for the same reasons).

### **Cognitive perspectives**

In cognitive perspectives, the metaphor that drives the conception of learning is that of *acquisition* [Mason, 2007]. If a student is not demonstrating a desired skill, it is because they do not yet possess fluent knowledge of that skill. Failure to demonstrate academic discourse would be considered due to students' forgetting specific language expectations, or students still acquiring meta-linguistic awareness of their own language behaviors. From a cognitive perspective, students' failure to demonstrate academic discourse in a particular instance would be the result of a cognitive mechanism such as forgetting, lack of meta-linguistic awareness, distraction, cognitive load, or lack of fluency in the language style [Siegel, 2010].

From a cognitive perspective, dialect integrated learning environments may support student achievement by improving their comprehension of instruction [Nieto, 2015], such as through reducing the cognitive load involved in comprehension [Sweller, 1988]. Edwards [Edwards et al., 2010] argues that each student may be at distinct stage of acquiring fluent bidialectal proficiency. Prior to achieving fluency in two language styles, code switching may entail needing to consciously override a dominant response, which may take substantial cognitive effort [Bettenhausen and Murnighan, 1991, Feldman, 1984, Bialystok, 2009]. Scholars have argued that for young students, continued use of AAE in the classroom, even after it has been clearly disallowed by their educator, may be due to students' lack of the *metalinguistic awareness* to differentiate between the two language systems [Edwards et al., 2010, Delpit, 1995]. Thus, from cognitive perspectives, dialect integrated environments may also support student achieving by improving their metalinguistic awareness, thus making it easier for them to comprehend and produce SAE during subsequent classroom instruction [Sweetland, 2008, Edwards et al., 2010].

### **Sociocultural perspectives**

Scholars from sociocultural perspectives would posit that *other* factors might *also* explain instances where students do not produce the expected academic discourse. Scholars of sociocultural perspectives highlight the historical contexts, attitudes, and identity orientations that are embedded into all interaction, including learning [Brown et al., 2005, Mason, 2007]. Within these perspectives, the metaphor driving learning is not acquisition but *participation* [Mason, 2007]. Brown et al. [2005] argues that educators should not presuppose that all students can take on the language requirements of the classroom free from social complication. From this perspective, language has the power to indicate social group affiliation and membership, even within (or, from some lenses, at the expense of) educational environments [Fishman, 1989]. For example, Godley and Minnici [2008] presents an ethnography on high school students wherein students reported they *could* speak SAE, but that they reserve its use for *important things* like job



interviews, not just classroom interactions. Similarly, Ochs [2005] identified that some students associate AAE with their *disidentification* from an academic identity.

In this thesis, we focus on how a design choice about the agent, its dialect, moderates the effectiveness of the same instructional paradigm across a group of students. In doing so, we situate this work in an emerging body of literature that has examined how aspects of a character’s design may carry social weight or activate social schemas that influence students’ learning performance. To better examine these explanations and test hypotheses, we turn to virtual agents.

## 2.2 Dialogic learning with (virtual) peers

Pedagogical agents have taken the form of tutors, coaches, and actors [Payr, 2003], experts, motivators and mentors [Baylor and Kim, 2005], learning companions [Kim et al., 2007], and lifelong learning partners [Chou et al., 2003]. Pedagogical agents have been shown to support students by delivering content and supporting both cognitive and metacognitive skills [Clarebout and Elen, 2007] through scaffolded guidance [Biswas et al., 2005]. In some cases, pedagogical agents have been able to monitor and adapt to students’ behaviors in order to individualize instruction [Sklar and Richards, 2010, Woo, 2009].

A common example of an effective pedagogical agent is AutoTutor, whose pedagogical strategies include the use of dialogue, feedback, corrective statements, hints, fill-in-the-blank questions, and requests for more information from the user [Graesser et al., 2008]. Pedagogical agents such as AutoTutor work, scholars argue, by guiding students through a productive dialogue where the technology serves the role of ensuring that the interaction is the sort of interaction that has been linked to learning. According to the authors, AutoTutor has been tested on over one thousand students, and produces learning gains of approximately one letter grade over traditional classroom learning environment outcomes. They also report that students who work with AutoTutor demonstrate better learning outcomes than those who work with *novice* human tutors, and almost reach the bar of *expert* tutors [VanLehn et al., 2007].

Scholars who examine the pedagogical effectiveness of virtual agents have identified that the way these systems are designed may impact how students perceive them, and that these factors ultimately moderate how effective the system is at promoting learning gains [Baylor and Kim, 2004, 2009, Clark and Choi, 2005, Domagk et al., 2010, Kim and Wei, 2011, Moreno and Flowerday, 2006, Ozogul et al., 2013, Schroeder et al., 2017, Veletsianos, 2010]. For this reason, pedagogical agents may be able to serve as *experimentation platforms* where we can evaluate the theories that have been identified as potentially important in supporting the needs of marginalized students.

Understanding how students may respond differently in different pedagogical agent learning environments may help us understand the mechanisms and moderating factors involved in human peer learning with greater nuance. Doing so may be possible in part because there is a large body of evidence documenting the social cues we take from virtual agents during interaction [Bailenson et al., 2003, Bickmore et al., 2005, Cassell et al., 2002, Gratch et al., 2007, Krämer, 2008].

When designing and researching virtual humans, researchers often focus investigations on increasing understanding of instructional design principles and effects, and how they influence cognitive outcomes for learners with different cognitive characteristics, in differently paced environments, or with content from different knowledge domains. Mayer [2014b] provides detailed overviews of many of these research-driven principles.

## **2.2.1 Explanations to date**

### **Cognitive perspectives**

Dialogic learning environments evoke opportunities for students to organize their own reasoning and recognize misconceptions through their ability to explain their ideas to another person. This idea has been corroborated through the evidence that students are more likely to learn from *giving* explanations than *receiving* them Webb [1989], Chi and VanLehn [1991], King [1990], Swing

and Peterson [1982]. There is additionally evidence that the type of dialogue being produced by students is proportional to their learning outcomes. This provides evidence to the idea that *something happens* through the process of dialogue generates knowledge or solidifies understanding in a way that may be demonstrated in subsequent assessments.

King [1989] refers to *thought provoking questions* as questions which ask students to make a claim and support that claim with reasoning. If students successfully respond to these dialogue opportunities with explanations, justifications, or other such *elaborated reasoning*, they are more likely to demonstrate positive learning outcomes [King, 1990, 1991, 1992, 1994, Chi and VanLehn, 1991]. Chan et al. [1992] found similar outcomes, identified that when children demonstrated *deep thinking responses*, similar to King's *elaborated reasoning*, there were directly observable links to students' outcome learning.

Scholars from cognitive perspectives additionally recognize the ways in which *thinking aloud* may have positive benefits in and of itself. For this reason, dialogic learning may be successful in part because of the opportunities it arises to support students in thinking aloud. Studies of individual thinking aloud have demonstrated that these interventions (though perhaps less natural than dialogic explanation) can successfully elicit students' idea elaborations in ways that help them reconceptualize their understanding that are positively related to post-test learning outcomes [Bargh and Schul, 1980, Ferrara et al., 1986].

### **Sociocultural explanations**

The idea that our social identities emerge through our language use has been argued using multiple frameworks and across multiple sources. Barton and Tan [2010] frame identity as what emerges when novices take on "position identities" made available in a particular practice [Holland et al., 1998]. Identity-in-practice develops through "how novices choose to accept, engage, resist, or ignore" social cues about appropriate dispositions and behaviors within a given situation [Barton and Tan, 2010]. However, we often don't notice these identity factors in others

unless it is *salient* to us, such as if it goes against our expectations or attitudes about what is appropriate in a given situation, or if we view it as personally validating and supportive to our own non-normative ways of being. When we deem someone *like ourselves* in a salient way, we may be more likely to feel a sense of belonging [Meyerhoff, 1996]), choose them as friends [Byrne, 1971], accommodate our speech patterns to be better aligned with their own language use [Abrams et al., 2002, Fought, 1999, Eckert, 2008], or emulate their behaviors [Berndt, 1999, Berndt and Keefe, 1996, Jones et al., 2012]. In learning contexts, these broad concepts of positive social environments have been associated with being a supportive environment for learning [Frisby and Martin, 2010, Madaio et al., 2016, Palinscar and Brown, 1984].

Often, peers are perceived to have either high or low *social capital*, an invisible imagined amount of social currency (such as popularity, or lack thereof). *Social impact theory* suggests that popular peers are especially influential and may have a disproportionate impact on how likely their own behaviors are to be emulated by others [Cillessen and Rose, 2005]. Conversely, *social misfit theory* posits that students who stray from the norms of social expectations, especially in a way that marks someone as low status, may promote both social or physical aggression [Dodge et al., 1990, Farmer and Rodkin, 1996]. This theory posits that while demonstrating the behaviors that are a mark of in-group behaviors may not consistently reward students with social capital and group acceptance, demonstrating *low-status behaviors* may result in students responding with *rejection* or *repulsion* [Wright et al., 1986].

AutoTutor, like many other technological and non-technological dialogic learning environments, are often in strong adherence to the cognitive elements underlying the benefits of dialogic learning. However, the developers of these systems often recognize the roles that emotions or affective states play throughout the learning process [Boekaerts, 2007, Calvo and D’Mello, 2011, DMello and Graesser, 2012, Pekrun and Linnenbrink-Garcia, 2012]. For example, a subsequent version of AutoTutor referred to as *Affective AutoTutor* was designed to be able to regulate students’ perceived negative emotional states such as boredom and frustration in order to increase

engagement and task persistence.

## 2.3 Virtual agents and identity

The *persona effect* is the name that has formally been given to the idea that the social cues which are exhibited by pedagogical agents can increase learners' motivations [Barlow et al., 1997], cognitive engagement [Johnson et al., 2000, Mayer et al., 2003], self-efficacy [Atkinson, 2002], and transfer achievement [Moreno et al., 2001]. This theory has been argued due to research that broadly shows that students learn a computer-based lesson more deeply when it is presented in a social environment than when it is presented through traditional text-and-graphics methods [Moreno et al., 2001]. Dismissing the argument that these environments were simply more *fun*, [Kim et al., 2007] report that they found that learners saw the agents as social models, expecting their agent to have a personality. For this reason, it has been argued that the design choices made in the creation of a virtual agent has the potential to *activate* these different social expectations [Haake and Gulz, 2008, Baylor and Kim, 2003, Veletsianos, 2010]. Despite this, [Schroeder and Gotch, 2015]'S review found that many agents were simply information delivery vehicles rather than facilitators of learning through complex pedagogical processes like demonstrating or scaffolding. It also demonstrated that agent design characteristics are often not considered for their potential sociocultural impacts.

Researchers have investigated topics such as the agent's physical appearance in relation to the context of the learning materials [Veletsianos, 2010], as well as their gender, age, or ethnicity [Johnson et al., 2013a, Kim and Wei, 2011, Moreno and Flowerday, 2006, Ozogul et al., 2013]. According to Veletsianos [2007], an agent's appearance could make a difference in how it is perceived, and thus influence learning. In a review done by [Heidig and Clarebout, 2011], it was concluded that the choice of agent design reflects the social perceptions such as competency and appeal, which in turn promote or hinder learning. This may be in part because the activation of social expectations is triggered by the visual and vocal characteristics of a virtual agent

[Haake and Gulz, 2008, Baylor and Kim, 2003, Veletsianos, 2010]. Agent design choices such as appearance, clothing, hairstyle, voice recording or generation process, sentence bank, and non-verbal behavior patterns may impact how users perceive the agent’s gender, intelligence, ethnic background, hobbies, and education level. These perceptions, in turn, are linked to differences in users’ perceptions of the agent’s task competency, emotional state, or trustworthiness, and other putative psychological states.

Domagk et al. [2010] conducted a study investigating the impact of learning with likeable, disliked, and neutral agents. In her study, she investigated how these different types of agents influenced learners’ motivation as well as their learning outcomes. Through two experiments, she found that while these social cues did not impact learners motivation, they did impact student learning. The students who worked with agents that had both an unappealing voice and appearance had lower transfer test scores than those in the other agent groups. Based on her findings, she concluded that “while differently appealing social cues might all trigger social responses, their effect may differ in the degree to which the learner is willing to actively engage in the learning process. Hence, unappealing social cues might even hinder learning [Domagk et al., 2010].

### **2.3.1 Virtual agents and dialect**

Having reviewed recent results in both dialect in the classroom and the implications of virtual agents in the classroom, we can now turn to their intersection. Kuhne et al. [2013], for example, presented an experimental study examining the effects of dialect regarding linguistic alignment, the likeability of the agent and the perception of the interaction based on whether the agent spoke *high German*, the prestige dialect in the study’s region, or *low German*, a stigmatized regional dialect. The study involved adults who were from the region where they presumably had been exposed to the low German dialect. The authors found that linguistic alignment occurred in both conditions. People who talked with the High German-speaking agent used fewer dialectical

words, those who talked to the dialect-speaking agent used more dialectical words than they had at pre-test. Kuhne et al. [2013], for example, report that the dialect-speaking agent was rated more likable than the High German-speaking agent, arguing that this is due, in part, to the *we-feeling* that is evoked when one interacts with an other that speaks with similar salient language features that someone associates with being an index of membership into one of their own perceived social groups [Aronson and Carlsmith, 1968, Bichel, 1973]. Kuhne et al. [2013] also found that those who talked with the High German-speaking agent linguistically aligned to a greater extent than people who talked to the dialect-speaking agent. They argue that this result is intuitive, “because in general dialect-speaking Germans are used to adapt to High German when speaking with Germans not from their dialectal region.” Interestingly, the authors also found that despite this improved *liking*, participants rated the interaction with the High German speaker as more positive overall. The authors hypothesized that the serious appearance of the agent didn’t fit to the rather chummy verbal behavior as which the Rhine-Ruhr dialect might have been perceived. In other words, there was a hypothesized mismatch between the agent’s dialect and the rest of the agent design characteristics or task context, similar to the mismatches referenced in the above section.

In a separate study, Astrid et al. [2010] demonstrated that speakers of a non-standard German dialect demonstrated rated a virtual agent as more likable when the technology spoke that same dialect, rather than standard German. Krenn et al. [2014] compared the social perceptions of an agent who spoke Austrian standard, colloquial Viennese, and dialectal Viennese. Agents who spoke dialectal Viennese, the most socially stigmatized of the three varieties, were perceived as sounding the least educated, but also having the best sense of humor. In free response follow-up questionnaires, the agent who spoke Standard German was most commonly assessed to be an academic who lives in the city, while the agent who spoke the stigmatized dialect was most commonly assessed to be a peasant who likes to go to the pub. Findings like this indicate that differences in user responses to an agent based on its dialect may not necessarily be due to the

dialect *itself*, but rather differences in the perceived *social categories* that are stereotyped to be *associated* with that dialect. In other words, it may be more accurate to say that, in this study, participants rated an agent more likable when that technology “sounded like someone they’d hang out with at a pub” than one who “sounded like an academic.”

A few recent studies have found that an agent’s voice might have impacts on a user’s experience with the technology. This has been referred to as *the voice effect* [Mayer, 2014a], and it posits that student learning will be improved when a standard-accented recorded human voice is used by an agent rather than a computer-generated voice. The majority of work on voice in this field has focused on comparisons between so-called “standard-accented” and computer-generated voices, finding that the human voice improve the users’ sense of trust, as well as increasing the amount of interaction users demonstrated with the system [Cherif and Lemoine, 2017]. Mayer et al. [2003] similarly found that students who watched videos narrated by a human voice rather than a computer-generated voice resulted in better retention and transfer of learning materials. In a follow-up study, Atkinson et al. [2005] replicated the experiment on 40 high school students and found similar effects for both immediate and delayed tests of knowledge transfer.

There has been substantially less research investigating “the voice effect” on *non-standard* voices, largely limited to cases I have identified earlier, where authors *have* investigated the impact of agent dialect have reported differences in users’ behaviors [Kuhne et al., 2013]. The authors of that work argue that this finding is in line with other work that has looked at speech accommodation with computer systems, including Branigan et al. [2010] and Brennan [1998], which both review a series of studies that find that humans demonstrate speech alignment in interaction with computers to the same rate (or a greater rate) than they do with other humans.



## Cultural Agency Theory

These findings are broadly in line with the theories posited by social agency theory [Louwerse et al., 2005, Mayer et al., 2003]. *Social agency theory* suggests that social cues in a multimedia message can prime the social conversation schema in learners in ways that may impact their interactions. Importantly however, social agency theory states that agents are able to *take on* the cultural or social factors that users associate with them based on their voice, appearance, behaviors, or so forth. For example, there is evidence that learners stereotype agents based on their appearance [Veletsianos, 2010]. This may have benefits, such as an increased potential for acting as a social role model if they are perceived to be within the students' same social group [Kim and Baylor, 2006, Rosenberg-Kima et al., 2008].

Arguing in support of these theories, Rosenberg-Kima et al. [2008] finds that a learner's interest and self efficacy in a domain task will be enhanced when they observe virtual models whose characteristics resembled them successfully perform a particular task. Their experiment showed that female college learner's interests and self-efficacies were positively influenced by the presence of a peer-like agent rather than an expert-seeming agent. Students of color were more likely to choose pedagogical agents that share the same ethnicity with them [Moreno and Flowerday, 2006], and college students preferentially chose agents whose gender and ethnicity matched their own [Kim and Wei, 2011].

In spite of this body of work, results with culturally aware technology have been mixed. Henderson (1990) began arguing for a more nuanced recognition of students' cultural and identity backgrounds in educational technology research more than 20 years ago. In her work, she argues that designers of educational technology may be particularly guilty of two types of mistakes in their systems. The first is a *deracialization* approach, where all aspects of the design of the system are those that align the system with the *unmarked* and *default* preconceptions of what a "*normal*" child will need. In these cases, "*normal*" is largely code for *child who behaves in ways that align with dominant societal expectations, and simultaneously values those*

*behaviors*. The second mistake is *essentialization*, where one stereotyped aspect of a cultural group is overemphasized to the exclusion of all other nuance. Interacting with a system where your own cultural background seems trivialized and stereotyped may be particularly insulting or distracting, or may cause the student to disidentify from the learning experience.

For example, Moreno and Flowerday (2006) evaluated how well students learned with virtual agents that either shared, or did not share, their gender and ethnicity. Half of their subjects were randomly assigned a same or different agent, and the other half were free to choose the agent of their choice. These authors found that students of color largely chose an agent who matched their own demographics from an initial image. However, they also reported that these students scored lower in learning achievement than those who were either assigned a white agent or voluntarily chose to work with a white agent. The authors concluded that this was because the social cues *interfered* with the students' ability to focus their attention on the learning material. They posit that the learners may have been focusing their attention on how the pedagogical agent represented them, and thus that working with a same-ethnicity agent had been harmful to their learning performances.

Looking at this example more deeply, as Moreno and Flowerday mention themselves, the authors changed the physical appearance of the characters, but there were no differences in language or behavior in each agent. Thus, what the authors describe as students' *distraction* by seeing how they were represented, may perhaps more accurately be described as students' disappointed recognition or *disidentification* with a learning process through which their identities were being trivialized, marginalized, and essentialized. The authors posit this explanation in their discussion, citing relevant work that posits that if the social cues provided in "same-ethnicity" pedagogical agents are not consistent with students expectations, their perceptions about the program are likely to hurt as well (Shneiderman, 1997). Further, the authors compare that there were differences in performance among students of color who chose to work with an agent of color than those who chose to work with a white agent. It may also be the case that there is relevant

social information between students who would or would not choose the same-ethnicity agent, such as how much they are aligning with that particular identity category in that instance.

Finally, I also point out here that student factors may moderate the effectiveness of an agent. For example, Sullins et al. [2009] randomly assigned participants to receive a narration on computer literacy by an agent presenting positive emotions by smiling and nodding or negative emotions by frowning and looking skeptical. They found that low ability students were *positively* impacted by an agent with an outwardly negative affect. In short, low ability students that received information along with an agent expressing negative affect performed significantly better than those in the positive affect group. One potential explanation for this might be that low ability students *themselves* did not feel positively about the learning experience, and thus were more willing to engage with a peer who *also* seemed to be doing the task begrudgingly, rather than one who seemed chipper about collaboration.



## Chapter 3

# Language annotation methods

In this chapter, I provide an overview of specific operationalized choices that span across studies presented in this work. The science domain that we study is held constant across interventions, aligned with age-appropriate expectations of students. Similarly, primary behavioral and attitudinal measures of interest are invariant across studies. Throughout this thesis, I present our findings about students' own language behaviors in this science domain and in terms of their attitudinal features. The measures presented here were broadly identified through the research of previous scholars, and then customized for our own purposes based on classroom observations, teacher interviews, and our abilities to reliably annotate for certain language features of interest.

The ways in which I have conceptualized science performance have shifted as I have become more familiar with the language behaviors consistently demonstrated among students in our populations, and as I've gained more nuanced hypotheses about students' language productions. For example, in initial conceptions of this work, I did not focus on *science discourse*, but rather what I had referred to as *school-ratified science arguments*, which were blocks of student reasoning that contained both what I now refer to as science discourse *and simultaneously* the presence of certain types of school-ratified evidence-based science reasoning. As I gained a greater theoretical background about the ways in which students' use of evidence-based reasoning may involve different processes than their discourse genre, I decided to separate out these two factors. At this

time, prior data was re-annotated to align with a more nuanced understanding of these language variables.

Similarly, our earlier analyses of this research aimed to investigate students' dialect use both in terms of their morphosyntactic features (e.g., grammatical structures) as well as their phonological features (e.g., accent and pronunciation). However, we ended up narrowing our investigation to exclusively morphosyntactic features, for two reasons. First, there was theoretical evidence that young students demonstrate a greater ability to demonstrate code-switching of their grammatical structures than their pronunciation, and that the majority of code-switching work with young students in education looked primarily at their production of Standard English morphosyntax. Second, we found it difficult to achieve inter-rater reliability on identification of students' phonological features. Doing so reliably, with objective boundaries between standard and non-standard pronunciation variants, often required the use of additional technologies that visualized the *wave forms* of students' language. The combination of these factors together resulted in our choice to focus exclusively on students' dialect use as defined by their morphosyntax.

In this chapter, I present the *final form* of how science discourse and dialect use were operationalized and annotated across each of the studies presented here. At the time of this thesis presentation, all annotation across each of the three projects was done using the same operationalizations of both science discourse and dialect use.

### **3.1 Pedagogical domain**

We situate our educational interventions in a science domain for several reasons. First, science is one of the academic domains in which we have historically failed to achieve equity for students of color Boykin and Noguera [2011], National Center for Education Statistics [2017]. Second, as science is a domain largely mediated through language and with specific expectations about language, it is a compelling domain in which to examine students' dialogic processes with a

virtual agent.

The educational interventions used in this research involved exposing students to *school-ratified science arguments*, using both *science discourse* and *evidence-based science reasoning*. We worked with teachers to identify age-appropriate science literacy expectations, using the Next Generation Science Standards (NGSS) as a guide. The NGSS are the product of a multi-state collaborative effort between the National Academy of Sciences, Achieve, the American Association for the Advancement of Science, and the National Science Teachers Association to create national teaching standards for science education. The goals of the NGSS are to standardize learning objectives as well as to foster and support interest in science, by building science-based skills in specific content areas and skills in “critical thinking and inquiry-based problem solving” [The Next Generation Science Standards For States, 2014]. The standards are specific to grade-level, and are divided into the topics of Life Science, Earth and Space Science, and Physical Science, each including sub-domains like *Interdependent relationships in ecosystems*, *Inheritance and variation of traits*, or *Forces and interactions*.

For each topic, there are specific performance expectations. For example, for the topic of *Interdependent relationships in ecosystems*, the following expectations are listed:

- (1) *Construct an argument that some animals form groups that help members survive*
- (2) *Analyze and interpret data from fossils to provide evidence of the organisms and the environments in which they lived long ago*
- (3) *Construct an argument with evidence that in a particular habitat some organisms can survive well, some survive less well, and some cannot survive at all*
- (4) *Make a claim about the merit of a solution to a problem caused when the environment changes and the types of plants and animals that live there may change*

We used these above science guidelines, with input from teachers, to design the specific content of the tasks students worked on with our virtual agents. We then applied our own experimental constraints towards these science activities. These constraints were enforced to help

ensure that the data produced from the intervention would allow us to answer our target questions as cleanly as possible, and included the task should be predominantly mediated through language (rather than object manipulation), *the tasks should not require specific domain content knowledge to complete successfully*, and *the tasks should allow opportunity for both everyday science language as well as school-ratified science discourse*. For each study in subsequent chapters, we will describe the specific tasks completed by students.

## 3.2 Language Measures

This work relies heavily on verbal annotation protocols (see [Chi, 1997] for an overview). For each student utterance, we annotated *science discourse*, *science reasoning*, and *task contribution*. Verbal annotations are not always easy to define due to false starts or links between ideas that were difficult to interpret. This is particularly true in annotating for student language, as doing so often requires the ability to interpret some amount of attending meaning from the students' speech; and furthermore, is more difficult in dialogic exchanges, where students and our virtual agent collaboratively built upon ideas together. Bereiter [Bereiter, 1994] called this "progressive discourse," where sharing, questioning, and revising of opinions leads to new understandings that are built up over the course of the dialogue, move by move, so that new meanings are consistently emerging [Wells, 1999]. While annotations were made at the utterance level, all quantitative results are normalized by duration rather than utterance count, and is therefore less sensitive to the definition of utterance breaks.

### 3.2.1 The language of science

In the science domain in particular, collaborative speech among students can be messy and generative [Edwards and Mercer, 1987], and science understanding is *talked into being* [Ash, 2004, Gallas, 1995, Green and Dixon, 1993]. Wenger [Wenger, 1998] argues that when looking at how



a particular exchange is either demonstrating an appropriate discourse (or not), it is important to look at the context of the full exchange in which an utterance is situated.

We annotated for each of the feature categories described in Figure 3.1 to quantify students' science language production. In this work, we separately define and annotate *science reasoning* and *science discourse*. Students could demonstrate school-ratified evidence within an everyday discourse genre, or use a science discourse genre for non-school ratified evidence types. In subsequent chapters, we primarily focus our results on students' demonstrations of science discourse, including its use with non-school ratified evidence. This was in part to allow us to keep a more specific focus on the bar of science discourse being a type of language style, without additionally confounding that with students' demonstrations of (or acquisitions of) school-ratified evidence. We also made this choice because it helped to control for differences in students' prior knowledge, whereby students who have stronger backgrounds in science might have better access to some types of domain evidence.

For all annotations described in this section, undergraduate students first achieved a Krippendorff's  $\alpha > 0.7$  on a subset of the data. All data was then split among students and each utterance was annotated by a single annotator. Utterances from the training subset were re-annotated by this same process to ensure consistency.

### **Science contributions and science reasoning**

Performative science language has been measured across a broad range of different types of measures including school-ratified epistemological structures [Ballenger, 1997, Warren et al., 2001], reasoning styles [Bazerman et al., 1988], and presentations of evidence Germann and Aram [1996], Jackson et al. [1993]. We first annotated for students' *science contributions*, where individual utterances were marked as having science content as long as there was *any* reference to the task. We then extend this definition to *science reasoning* when students used this content from the task as part of the application or justification of an objective claim.

Type	Category	Description	Example
Evidence type	School-ratified	Evidence based in objective source such as recall from domain-ratified prior knowledge or observable data in the task	I think that <u>he can eat</u> the bunnies because he has those monster <u>teeth</u> that could destroy anybody I think he's a <u>good climber</u> because those <u>long arms</u> are gonna be able to climb and climb and swing and swing, all up anywhere he wants to go
	Non school-ratified	Evidence based in personal preference, opinion, or anthropomorphic assertions	I don't think the creatures like to be near him because they <u>don't trust</u> him I think he probably sleeps in the softest nest underground so he <u>doesn't feel</u> sad
Discourse genre	Science	Reasoning structures that make use of epistemological connections dominant in Western science communities	I think the creature <u>can probably eat</u> worms because <u>his hands look like</u> they can pick up stuff really easily <u>If the creature is able to camouflage, then</u> I think he goes out at night <u>so</u> he's safer
	Everyday	Reasoning structures that make use of anthropomorphic descriptions, everyday analogy terms, or affective preferences	I think the creature is <u>always saying like</u> "leave me alone, leave me alone lions! I'm bigger than you and I can take you!" I think <u>he finds</u> the best climbing tree to sleep in <u>and then</u> he's able to not let anyone <u>defeat</u> him
Science contributions	Content	Utterances that contribute content about the activity regardless of reasoning style, evidence type, discourse genre, or so forth. Often includes observations	Don't you just think the creature is so weird and ugly? He has the biggest frog mouth I've ever seen on something that's not a frog
	Non content	Utterances that do not contribute science content, including meta-commentary about the task, social exchanges, or questions about the agent's abilities	How much longer until this is over? Do you like living in the computer?

Figure 3.1: Broad feature annotation scheme for students' science language production.

## Science discourse

If we here define science reasoning as the application of information from a science task to justify a claim, then science discourse is defined in parallel as the language *style* used to present that reasoning. Some scholars argue that science discourse is less *a specific checklist of language moves* and rather *the set of language patterns that make one "sound like a scientist"* [Lemke, 1990, Lampert, 1990, Yackel et al., 1990, Halliday and Martin, 1993, Heath, 1983, Norris and Phillips, 2003]. For this reason, some scholars have avoided fully quantitative metrics of this

variable. For example, Brown [Brown et al., 2017b] and Mercer [Mercer and Dawes, 2008] use discourse analysis techniques to provide a picture of *how* students are talking within an exchange.

In our operationalization of this variable, we took science discourse to be the demonstration of the *grammar of science*. Based off of prior work on students' epistemological reasoning structures [Ballenger, 1997], field work with teachers, and the Next Generation Science Standards, we defined science discourse as *connections between a claim and provided evidence that relies on causal explanations using dominant patterns of Western reasoning*. These largely included structures such as "*I think [x] because [y]*", "*If [x], then [y]*", "*[x], and so [y]*," and other linguistic structures that capture causal reasoning.

## Science arguments

We also report students' use of *science arguments*. Science arguments were defined here as utterances that involved both science reasoning and science discourse together. Science arguments would be the only student utterances largely recognized as fully successful science literacy. In all studies, our virtual agents exclusively demonstrated school-ratified science arguments in both conditions. In other words, all science reasoning from virtual agents was evidence-based, and it was always presented in science discourse.

However, due to strong theoretical motivation to believe that science discourse would be impacted more easily than science reasoning (because it required no additional prior knowledge over the "*grammatical construction*" of the sentence), in all studies we first report on students' science discourse as our primary hypothesized finding. We largely make this distinction due to limited experimental power, and a consistent desire to reduce the number of primary analyses run in each experiment.

### 3.2.2 Dialect density

Standard American English (SAE) and African American English (AAE) were previously defined in the introductory chapter. Rickford [1999] provides a list of phonological, morphological, and syntactic features that are common to AAE. Green [2002] supplements this with an even more detailed account of linguistic features typical of AAE speakers, including an in-depth specifications of lexical, semantic, syntactic, morphosyntactic, and phonological properties. In this work, we focus on the morphosyntactic, rather than phonological, features of AAE and SAE. This is both because they are easier to annotate, as well as because they are easier for children to use intentionally Craig and Washington [2006].

As non-standard English dialect use is so controversial, we began this research by gaining additional insight about how students use dialect *in-vivo*, and how teachers responded to this dialect use. Our classroom observations took place across four local charter schools, where we focused our observations on eight elementary school teachers. The educators we worked with were representative of their school and national demographic statistics of public school teachers ?. These schools report African American student populations of between 65% to 100% and provide between 75% to 99% of students with free or reduced lunch. These features were identified through initial classroom observations, were iterated over the course of this work based on the observed frequency among our participants, and are described in Table 3.1. The presented variants are the most common AAE features observed in our sample of 2nd - 4th grade students in Pittsburgh, and were finalized in our analyses following our data collection efforts.

Utterances were annotated for presence of these features in either SAE or AAE variation. Unlike other annotations described in this chapter, all annotators of grammatical features were senior undergraduates pursuing degrees in Linguistics. With this constraint, inter-rater reliability for all eight features was reliable ( $\alpha > 0.7$ ); without specialized annotators, annotations were unreliable. Once reliability was attained, all data was distributed among these specialized annotators and received one annotation, including the subset initially used for training.

Table 3.1: Morpho-syntactic SAE and AAE variants annotated

Feature	Definition	SAE variant	AAE variant
Copula	The omission of the conjugated form of a “to be verb, joining a subject and a predicate	He is a predator	He a predator
Auxiliary	The omission of auxiliary verbs, often called “helping verbs (e.g., can/could, have/has, do/did, etc.)	How do you think he eats?	How you think he eats?
Subject verb	The presence of a subject and corresponding verb that differ in grammatical marking for number and / or person	He looks fast	He look fast
Past tense	The omission of past tense marking (-ed) on verbs, or adding an -ed to an already past-tense word	Last time you said he was a carnivore	Last time you say he was a carnivore
Possessive	The omission of inflectional -s to indicate possession	The creatures spikes arent that sharp	The creature spikes arent that sharp
Plurality	The omission of -s when indicating the plural form of a noun where the SAE variant marks -s	He has all those spikes	He has all those spike
Negation	The presence of a double-negation or use of the word aint to mark negation	He doesnt have any fur	He aint got no fur

One of the most prevalent ways of quantifying students’ use of either SAE or non-standard variants such as AAE is the Dialect Density Measure (DDM) [Craig et al., 1998]. The initial use of this instrument was to gauge a speaker’s composite use of AAE. This methodology involves annotating the number of identified AAE features, and dividing by a normalization token, such as number of utterances or number of words. In the work presented here, we divide AAE tokens by total count of identified grammatical features from the set listed in Table 3.1. This was an important decision in this work, as more complicated reasoning structures generally involved more grammatical features. Dividing by number of utterances, or even words, could have conflated sentence complexity with AAE use. Here, our DDM variable results in a composite score of the percentage of grammatical features produced in AAE. By comparing students’ DDM scores across conditions, we can evaluate the impact of a virtual agent’s dialect on students’ own dialect production. By comparing students’ DDM scores within conditions, we can evaluate the level to which students demonstrate code switching.



## Chapter 4

### Study 1: Dialect as an elicitation technique

The goals of this initial experiment were to understand the immediate impacts of agent dialect pattern on students' language production and social perceptions. There were three primary goals of this study. The first was to perform a *manipulation check* on our instantiations of dialect and students' science demonstrations as a way to refine our understandings of these variables moving forward. As our primary questions throughout this work were built around understanding *why* bidialectal learning experiences seem to have benefits for students, this initial study was to assess to see if this was in fact a replicable result in our own conceptions of these variables. This study also served as an opportunity for us to collect samples of students' science production we could use to build a virtual agent.

For these reasons, this study was done prior to us fully building the interactive virtual peer we used for our next two experiments. We used a non-interactive character ("Jamie") that was presented to students as the pre-recorded speech of another student from another school, sort of like a pen pal. This experimental design choice allowed us to control for what students heard. While our next two studies using the virtual peer had introduced inherent confounds due to the nature of dialogue, this first experiment used non-interactive, pre-recorded speech such that all students heard *identical* content.

In doing so, this study was the first experimental manipulation of learning environment di-

alect on students’ resulting science language behaviors. However, there was a foundation of research (both between human interlocutors and between a human and a virtual agent) that led us to hypothesize there *would* be an impact of agent dialect. *The voice effect* refers to the phenomenon that demonstrates that the voice of an agent seems to impact students’ interactions with that agent, though this has largely referred to human-esque or machine-generated agent voices [Mayer, 2014a, Cherif and Lemoine, 2017]. At the same time, dialect differences both in humans and computers can impact someone’s social perceptions of an other [Giles and Marlow, 2011, Kuhne et al., 2013]. Kuhne [Kuhne et al., 2013] and von der Putten [Astrid et al., 2010] both identified that bidialectal human interlocutors accommodated their own dialect styles to the dialect used by a virtual agent, in these cases either Standard German or Dialectal German.

## 4.1 Study overview

In this study, we used a non-interactive agent named *Jamie* that served as a worked example of science discourse [Atkinson et al., 2000]. In a worked example, an “expert takes a well-formed problem and displays it as a model for learners in terms of how that problem is approached, thought about, worked over, and solved. The worked example is meant to model for newcomers how an expert thinks, values, and acts in a given domain. For this initial experiment, we showed all students identical pre-recorded examples of science discourse in either SAE or AAE to minimize confounds.

In this study, students were told that they would be working with an assigned partner from another school *just like theirs*, and that they would be exchanging speech samples with their partner “*kind of like a pen pal*.” In reality, all students were assigned identical partners (Jamie) that were different only in dialect use. Students were shown identical cartoon images of their partner, and were told that their partner designed their avatar to look like them. Students were told that at the end of the class period, they would be able to design an avatar of themselves to send back to Jamie, along with their recordings. Jamie’s speech was pre-recorded by a bidialectal



African American confederate.

The experiment involved both a *social introduction task* and a *science presentation task*. Though the *content* was identical, it was provided to student using one of three dialect patterns: monodialectal SAE, monodialectal AAE, or bidialectal. In the bidialectal condition, students' heard a social introduction in AAE, but the science example was in SAE. Thus, in this instantiation of our bidialectal condition, students exclusively heard the agent talk about science in SAE, using the same recording as the agent from the monodialectal SAE condition. Students were asked to provide their own social introduction after hearing Jamie's social introduction, as well as their own science presentations before and after hearing Jamie's science example.

This experiment also served to identify whether students' science production would be impacted by an agent who spoke AAE in the social task but SAE in the science task, or if there would only be differences in students' language production when the agent spoke AAE in the science task. This would elucidate whether or not hearing an agent speak AAE *at any point* would impact students' science production, or if there would only be benefits when the agent used AAE during the science task.

Because of the brief and non-interactive design of this experimental paradigm, we expected any differences observed here in students' language use to be better attributed to *imitation* than *learning*. For example, in sociolinguistics research, one measure of assessing students' linguistic competence is by altering the social context of a dialogue and evaluating differences in students' subsequent language production. For example, when Kindergarten students are asked to act out a hospital visit with dolls, they are more likely to *hedge requests* when they are pretending to be a nurse, and issue *direct orders* when they are pretending to be a doctor. Students may demonstrate sociolinguistic variation based on the social context in which they are speaking, such as toddlers speaking more slowly to adults with unfamiliar accents [van Heugten et al., 2015], or consistently altering their use of discourse markers ("because" vs. "so") based on the genre of the conversation [Sprott, 1993].

Relevantly to the work presented here, there is also evidence that students' perceptions of their interlocutor might impact their production of certain language features. For example, Alim [Alim, 2009] identified that high school students adjusted their density of AAE use based on their familiarity with an interlocutor, or the level to which they perceived their interlocutor would have knowledge of hip-hop. Similarly, Eckert [Eckert, 2000] argues that as students approach adolescence, students begin to place a premium on the language styles that earn them social access among their peers. Ogbu [Ogbu, 1999] and Godley ([Godley and Escher, 2012] both provide qualitative evidence that African American high school students may *choose* not to use Standard English in some contexts where doing so would challenge their personal sense of identity. Among elementary-aged students, there has been no experimental work to my knowledge that aims to understand if the dialect used by an interlocutor impacts their own subsequent language use - whether these differences are attributed to learning, imitation, or sociolinguistic elicitation.

Though it has been limited, learning from observing actions or behavior has been linked to improved student learning in some instances. For example, there is some evidence that one may be more likely to *pick up* aggressive behaviors after witnessing these behaviors in an other [Bandura, 1969, Bandura and Cervone, 1986]. Some of the social theories of learning reviewed in chapter 2 provide additional theoretical explanations for these *imitative behaviors*, such as peer imitation being more likely if a student feels that the peer is similar to them, or is someone who has increased social capital [Berndt, 1999, Berndt and Keefe, 1996, Jones et al., 2012, Cillessen and Rose, 2005].

Most of the research on peer learning demonstrates that learning is most likely to be influenced by how *actively* a student is participating in the interaction. In a brief non-interactive worked example, there were no opportunities for students to directly interact with the science model being presented. Rummel et al [Rummel et al., 2009] demonstrate evidence that it *is* possible to learn watching an example of someone else learn (what she referred to as vicarious peer tutoring), though she found that watching these tutoring sessions collaboratively with an other

improved the effectiveness of the model at promoting learning.

For these reasons, we expected any differences in students' language use in this study to be due to *imitation* or *elicitation* rather than learning. Though these distinctions are hard to separate, we expected that a brief, non-interactive worked example would likely not be able to promote students' improved acquisition of these language styles (or, if so, that they appear to be incredibly easy to teach). Instead, we expected that any differences in students' own language use following the model presented by the agent would be the result of them imitating salient aspects of Jamie's language, or perhaps Jamie's language creating a sociolinguistic contexts that elicited students' use of language features they were already capable of producing. If our results revealed differences in students' language use based on the agent's dialect, it might mean that agent dialect impacts the level to which students' accommodate their own speech to the agent's (either consciously or otherwise).

#### **4.1.1 Research questions**

We used this design to be able to answer the following questions:

- (1) Does the dialect (SAE or AAE) in which an imagined "peer" provides a science example impact students' subsequent science discourse? We hypothesized that students who heard an agent use AAE at any point in the experiment would demonstrate more science discourse than students who were assigned a monodialectal SAE agent.
- (2) Are there differences in students' own dialect based on whether the "peer" models monodialectal AAE or bidialectalism? We hypothesized that if students had fluency in both SAE and AAE, they would be more likely to code switch into SAE after seeing an example of code switching from the code switching agent.
- (3) Does this peer dialect impact students' social perceptions of the character? We hypothesized that students with the AAE-integrated conditions would self-report that

their “peer” partners were more likable than the monodialectal SAE agent, but also less intelligent, in line with other research on dialect and social perceptions (Giles, 2001; Kuhne, 2014).

We argued that the brief nature of this intervention language exposure would demonstrate which of students’ own language features were sensitive to agent dialect. Because we modeled a brief, non-interactive worked example of a very short period of time, we did not necessarily expect that students would *learn anything new* from the agent within traditional conceptions of learning as a complex cognitive process. Most accounts of how students learn from models focus on the importance of active participation and co-construction as a way to generate and integrate new knowledge [King, 1992, Chi et al., 1989]. This design did not allow substantial opportunities for this type of interactive cognitive work.

Instead, we expected that if students *were* sensitive to agent dialect, condition differences would be due to *elicitation* or *imitation*, rather than *teaching*. In other words, we expected any differences in this study to be due to differences in effort, motivation, engagement, desire to emulate (or distance) from the agent, or unconscious accommodation of language production, rather than complex processes of learning.

## 4.2 Methods

This study took place within the science classrooms of a local, Afro-centric charter school with lottery admittance that serves students kindergarten through 5th grade. The school demonstrates 100% African American student population, with 99% of students qualifying for free or reduced-cost lunch. This school served as one of our primary partner sites during our early qualitative work which is described in Chapter 5. This study took place approximately three months into that qualitative work, during which we had sat in on several of the students’ science classes. At the time of this study, we had not had any substantial interaction with any of the students directly. When our classroom observations had begun months before, we had been introduced by

their science teacher as “people from Carnegie Mellon who are here because they want to help figure out how to make science class more fun, like with computers and stuff.”

The study described here took place during both normally-scheduled 3rd grade science classes in students’ typical classroom. So as to not influence students’ behavior during the intervention, the science teacher left at the beginning of the class period with the students who did not return signed parental consent forms. Of the 45 total students enrolled in 3rd grade, 29 had returned signed forms and participated in our intervention. Technology failure rendered the data of six students unusable, leaving us with data from 23 participants. Students all completed the intervention at the same time on individual computers, and were arranged to minimize the proximity to others. The full intervention was completed by students within their 50-minute science period.

At the beginning of the class period, we told students that we were working on a technology that would let students from different science classes send messages to each other, “sort of like pen pals.” We told students that we were working with a few other schools *just like theirs*, and that today they would *hear messages from a partner*, and record messages that we would share with the partner later on. We used this manner of deception to minimize the potential confound of students’ beliefs about a technology using non-standard English. To our knowledge, no students outwardly demonstrated that they believed their partner was not actually a human peer. We refer to this paradigm as a *distant peer* paradigm.

Students were told that they were each randomly assigned a partner, and that their partner had made an avatar to represent themselves. Students were told that they would get to make their own avatar at the end of the activity before we sent their messages along to their partner. They were told that first they would listen to an introduction that their partner recorded for them, and that they would be able to record an introduction themselves. They were told that we would do a science activity with their partners after everyone had introduced themselves.

The only difference in Jamie between conditions was Jamie’s dialect pattern. Every student heard two examples of Jamie’s speech, in the form of four-minute monologues. The first example

was a *social introduction*, and the other was an answer to a science prompt similar to the one students would be asked to respond to both before and after hearing Jamie’s example. See figure [x] for a representation of this procedure.

In the **Standard American English (SAE)** condition, Jamie used SAE during both the social task and the science task.

In the **African American English (AAE)** condition, Jamie used AAE during both the social task and the science task.

In the **Code-Switching (CS)** condition, Jamie used AAE during the social task and SAE during the science task.

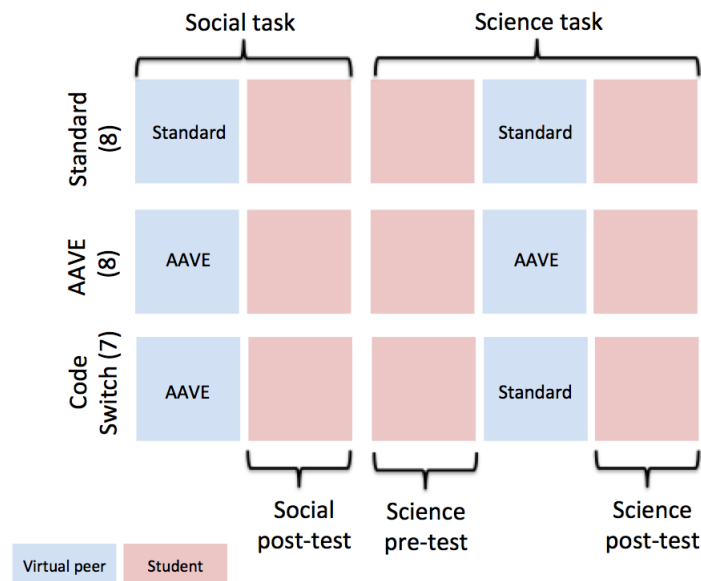


Figure 4.1: Procedure for the distant peer study across three conditions

### 4.3 Science activity

The science activity used in this initial analysis was designed to meet expected goals set by third grade learning standards. In each instance of this task, students were given two laminated prints of identical environments, each one with a different imaginary creature shown inside. Students

were asked to come up with as many hypotheses as they could for which creature they thought would be a better fit for that environment, and why. Students were shown one set of images for the pre-test, and another isomorphic set at post-test. The order of the image sets between pre- and post- were counter-balanced across students.

This was our first iteration of a potential task for students to complete with our eventual virtual peer. In this way, this experiment also served as a way for us to collect corpus data of what sorts of language students demonstrated in reasoning about ecological principles. We used this to refine our conception of age-appropriate science discourse and iterate on our understanding of students' science language production in line with our research questions.

Students were given one minute to look at their pictures before they began recording their answers. Once students confirmed that they did not have any questions, students' individual four-minute timers were started. After four minutes, students were told that they were now going to hear an example that their partner did last week. Jamie's science example was also a four-minute monologue, and included eight instances of science reasoning. More detail about the science content in this task is provided in section 4.x. Depending on their condition, students heard Jamie's science example in either SAE or AAE.

Following the example, students were asked to do the science activity one last time. They were again told that their recording would be shown to their partner, similar to how they just heard Jamie's recording. As described in chapter 4, students' recordings were annotated for science reasoning, science discourse, and AAE dialect density to determine if Jamie's dialect impacted any of these three features in students' post-test performance.

The creature set used in Jamie's example were designed to demonstrate features that were not directly present in the students' image sets. This was done to minimize students' ability to simply repeat domain words or phrases directly from Jamie in their own post-test recordings. Students' task sets displayed creatures living on land environments, and Jamie's task set displayed creatures living in an aquatic environment. Each example of science reasoning that Jamie



Figure 4.2: Science task images used in dialect elicitation study. Sets A and B were counter-balanced across pre- and post-test.

demonstrated was specific to this aquatic environment which students' did not have in their own task environments.

### 4.3.1 Agent science model

#### Visual appearance

Jamie's visual appearance was determined through a previous in-lab pilot study. We wanted to ensure that the avatar was gender-ambiguous to reduce the potential for gender-based confounds. In this pilot, students were shown six different characters, and were asked to write a one-sentence description of each one using the form "she is..." or "he is...". We looked at students' pronoun choice for each character to determine whether they perceived it to be male or female. The avatar chosen for use in the study was the one most commonly perceived to be the same gender as the child participant. Five of five boys perceived this avatar to be male, and three of five girls perceived the avatar to be female. (As an interesting aside, the only avatar that fared better among



girls was identified with feminine pronouns four of five times, but was not identified as male by any of the male participants).

### **Dialect use**

Jamie was voiced by a local, bidialectal African American woman, who produced an SAE and AAE version of both the social content and the science content. The content was initially written in SAE, and then translated to AAE with the help of our voice talent. This translation was done alongside the bidialectal voice actress, and was based on the common AAE features identified by prior scholars [Renn, 2010, ?] and our classroom observations. The voice actress worked with us to ensure the AAE speech sounded natural, and that the AAE features were used authentically. Jamie's dialect use was set using the Dialect Density Measure (DDM). The target percentage of DDM was based upon previous calculations of DDM among this age group within the literature. With a target of 8-9%, the DDM in the AAE version of our social task recording was 8.6% and the DDM in the AAE version of the science task recording was 8.2%.

### **Social introduction**

To inspire the content of Jamie's social introduction, we had asked the students in our pilot study to tell us about their families, their friends, what movies they liked, what they did in school, etc. We used the most common themes in Jamie's social script. The four minute social introduction provided by Jamie included information like how many siblings she has (*three*), her favorite movie (*Madagascar*), and her favorite sport (*basketball*). She shared that her cousins used to live nearby but that they're moving soon, and that she's going to miss them. She shared how she felt about math class (*it's okay*) and recess (*it's the best*).

Prosodic analyses determined that there was no significant difference in the volume, speed, or pitch fluctuation of these recordings to help us ensure that we were only manipulating the presence of dialect, and not other audio variables.

## **Science task**

Jamie's four-minute science example included 35 total utterances, eight instances of school-ratified science reasoning, and twelve instances of school-ratified science discourse. Because all instances of reasoning occurred with science discourse, this means that Jamie also demonstrated eight total school-ratified science arguments, with a science discourse density of 1.5. In other words, there was an average of 1.5 instances of science discourse per instance of students' science reasoning.

The other utterances included other on-topic science content including observations (it looks like the creature has gills), comparisons (one creature looks like it can stand up on both legs, but the other one looks like it can only swim), and questions (I wonder which one is more dangerous...). Each of Jamie's instances of science reasoning involved science discourse. Jamie did not provide any instances of everyday reasoning.

## **4.4 Results**

### **4.4.1 Science discourse**

We measured students' demonstration of science discourse during pre-test and post-test to understand if the dialect in which Jamie demonstrated science discourse impacted the likelihood that students demonstrate this discourse style at post-test. We used one-way ANOVAs to measure differences among the three conditions.

There was no significant difference in students' use of science discourse at pre-test by condition, with students on average demonstrating 2.77 instances ( $SD = 2.97$ ). A within-subjects t-test across all conditions demonstrated that, on average, students produced more science discourse after hearing Jamie's example, regardless of dialect ( $t(25) = -2.91, p < .01$ ). Across conditions, students demonstrated an average of 4.85 instances of science discourse at post-test ( $SD = 3.0$ ).

A one-way ANOVA on students' science discourse post-tests revealed that condition signif-

icantly impacted the extent to which students improved from pre-test ( $f(2, 23) = 6.85, p < .01$ ). A Tukey post-hoc test revealed that there were significantly more instances of science discourse among students who heard Jamie's science example in AAE ( $M = 7.11, SD = 2.32$ ) than those who exclusively worked with an SAE-speaking Jamie ( $M = 2.6, SD = 1.41, p < .01$ ). Students who worked with the code-switching agent (and thus heard the science example in SAE) had science discourse post-test scores between the other groups, not significantly different from either ( $M = 4.56, SD = 3.32$ ).

Within-subjects gain was examined by condition with a paired-subjects t-test by condition. This test revealed that while students in the AAE-only condition demonstrated a significant gain in science discourse from pre-test to post-test ( $t(8) = -7.16, p < .0001$ ), the other two conditions did not ( $p > .05$ ).

We additionally assessed students' use of *everyday* discourse as an alternative way of supporting science arguments. While an argument made with *science* discourse might be "The creature looks like a predator because it has sharp claws," a similar argument made with *narrative* discourse might be "Whenever I see something with claws like that, I know I have to run run run because it's gonna be definitely something that eats everyone around it."

As with science discourse, narrative discourse was coded independent to science reasoning. There was no significant difference in students' production of narrative discourse by condition at either pre-test or post-test. At pre-test, there were an average of 2.2 instances of narrative discourse ( $SD = 2.15$ ), and at post-test, an average of 1.5 instances ( $SD = 1.48$ ).

#### 4.4.2 Science arguments

As outlined in the Chapter 3, science arguments are instances of school-ratified science reasoning that were additionally communicated using science discourse. Though we were primarily targeting students' use of science discourse as our primary outcome variable under investigation, we additionally present here our findings regarding science argumentation.

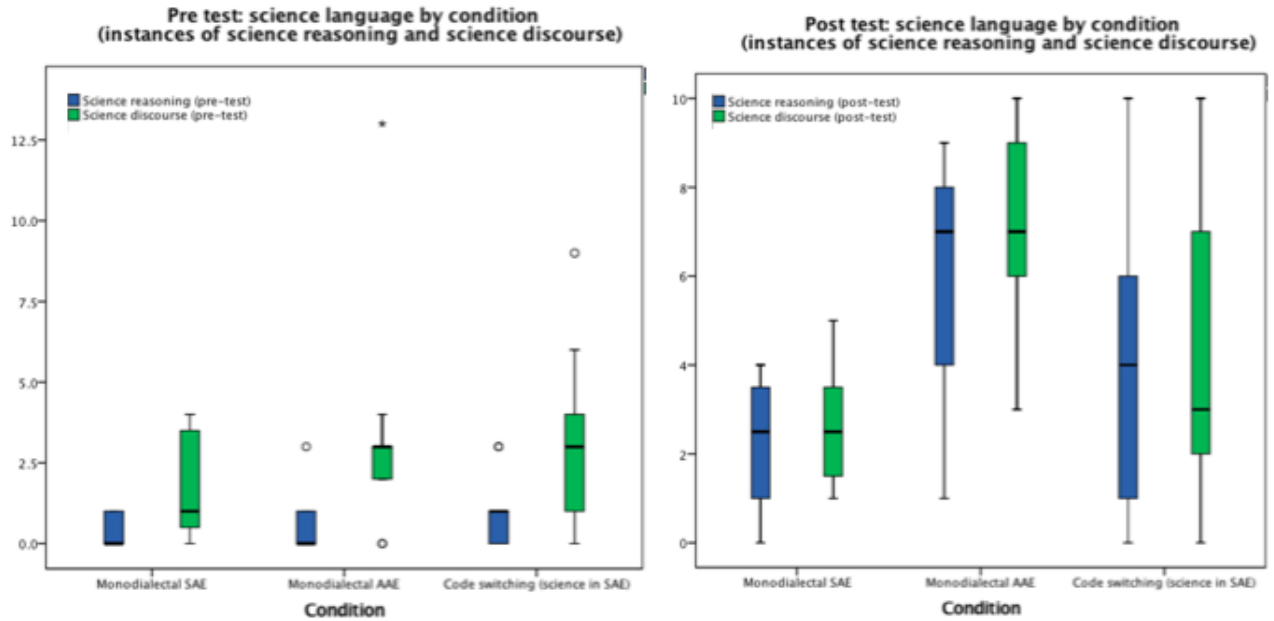


Figure 4.3: Pre and post test assessments of students' science reasoning and science discourse by condition

We first look directly at students' science reasoning regardless of science discourse. There was no significant difference in students' use of science reasoning at pre-test by condition, with students on average demonstrating an average of less than one instance of reasoning at pre-test ( $M = .69$  instances,  $SD = .97$ ). A within-subjects t-test across all conditions demonstrated that, on average, students produced more science reasoning after hearing Jamie's example, regardless of dialect ( $t(25) = -5.48$ ,  $p < .0001$ ). Across conditions, students demonstrated an average of 4.08 instances of science reasoning at post-test ( $SD = 3.03$ ).

A one-way ANOVA on students' science discourse post-tests revealed that condition significantly impacted the level to which students improved from pre-test ( $f(2, 23) = 4.14$ ,  $p < .05$ ). A Tukey post-hoc test revealed that there were significantly more instances of science discourse among students who heard Jamie's science example in AAE ( $M = 6$ ,  $SD = 2.55$ ) than those who exclusively worked with an SAE-speaking Jamie ( $M = 2.25$ ,  $SD = 1.59$ ),  $p < .05$ . Students who worked with the code-switching agent (and thus heard the science example in SAE) performed

between both groups, not significantly different from either ( $M = 3.78$ ,  $SD = 3.5$ ).

Within-subjects gain was examined by condition with a paired-subjects t-test by condition. This test revealed that there were significant gains in science reasoning among students in the AAE condition ( $t(8) = -7.16$ ,  $p < .0001$ ) and the SAE condition ( $t(7) = -2.81$ ,  $p < .05$ ), though not among those with the code-switching agent ( $p > .05$ ).

While students in both the monodialectal SAE and monodialectal AAE conditions demonstrated increases in science reasoning from pre-test, only the monodialectal AAE condition demonstrated within-subjects increases in their science discourse. Additionally, students who heard the science example in AAE demonstrated significantly more instances of science reasoning and science discourse than those who worked with the SAE-only character. We examined if this was due to a difference in *amount* of speech, and determined that there was no significant difference in word production at post-test by condition. In fact, while students with the AAE-only agent demonstrated an increase of 45 words from pre-test to post-test ( $SD = 84.9$ ), those with the SAE-only agent demonstrated a generally larger, but also non-significant, increase of 88 words ( $SD = 122.7$ ).

To investigate the relationship between science reasoning and science discourse, we examined their co-occurrence as a measure of their *science argumentation*. While our results demonstrated that students with the monodialectal AAE character demonstrated more instances of science reasoning, were they additionally applying science discourse at a greater *rate* than the students with the SAE character?

To do this analysis, we assessed the percentage of students' school-ratified science reasoning that additionally contained science discourse - in other words, the percentage of arguments involving school-ratified reasoning that were *also* produced using science discourse. Science arguments were marked as having science discourse if at least one instance of science discourse occurred within the set of utterances composing the reasoning. Arguments were marked as having science discourse even if they co-occurred with instances of narrative discourse.

For example, the following lines were produced by a student working with the monodialectal SAE agent.

“And then I’d tell her that the right ones were gonna get ’em the fastest because they have those spi- spikes that can protect ’em from the other- the predators that are gonna be chase-chase-chasin’ ’em around.”

This instance was marked as having school-ratified reasoning because of the use of school-ratified evidence (*the spikes that can protect from predators*). It was also marked as having both science discourse (*the right ones are fastest because they have spikes for protection from predators*) as well as everyday discourse (*the predators are gonna be chasin’ ’em around*). (Note, the use of *chase-chase-chasin’* here was unanimously agreed upon by annotators to be for narrative effect, rather than the result of false-starts or stuttering).

Paired-samples t-tests revealed that across conditions, students demonstrated a significant increase in their instances of science argumentation, from an average of .58 at pre-test (SD = .76) to an average of 2.7 at post-test (SD = 2.3),  $t(25) = -4.1$ ,  $p < .0001$ . On average, those who exclusively heard the agent speak in SAE demonstrated 1.3 additional science arguments from pre-test to post-test (SD = 1.8), while those who exclusively heard the agent speak in AAE demonstrated 3.3 additional science arguments at post-test (SD = 2.6). Similar to the previous results, those who worked with the code switching agent (and heard social talk in AAE but the science in SAE) performed between the two groups, gaining 1.7 additional science arguments at post-test (SD = 3.0). A one-way ANOVA revealed no significant differences between these gains by condition ( $p > .05$ ).

### **Science argument density**

Because our prior results revealed that students with the monodialectal AAE agent demonstrated more instances of science reasoning than those with the monodialectal SAE agent, we calculated the *density* of science argumentation. We expected that if students were *learning* from Jamie,

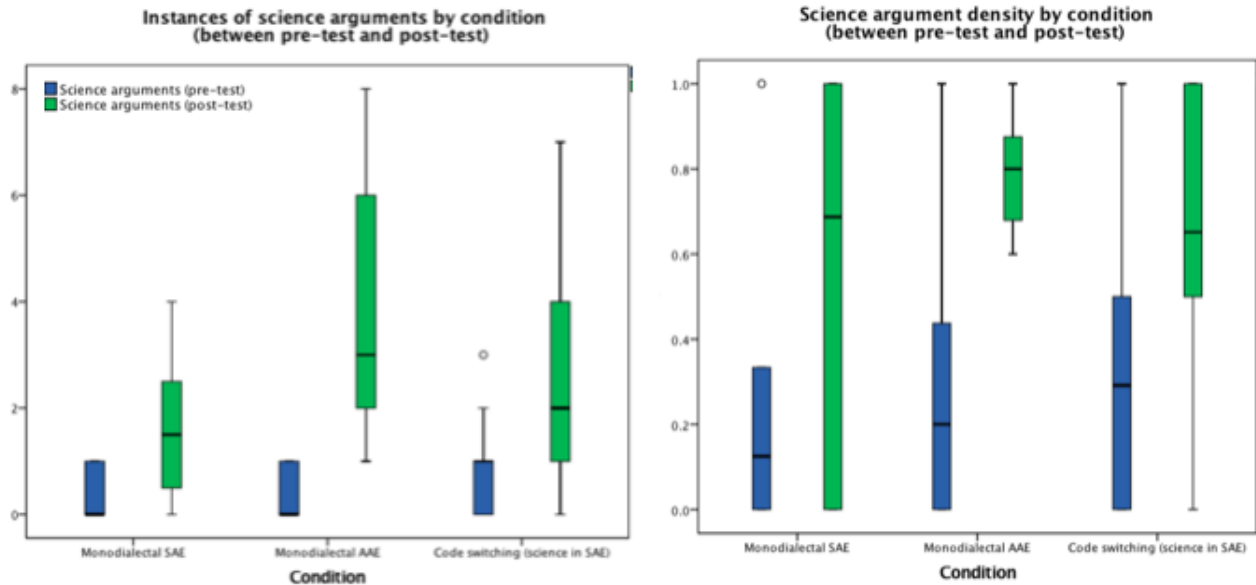


Figure 4.4: Instances of science arguments (left), and percentage of arguments involving science discourse (right)

they would demonstrate an increased percentage of reasoning that included science discourse than the rate of science discourse use at pre-test. If they were simply demonstrating *more science* with Jamie (but perhaps not categorically *different science* with Jamie), we would expect there to be no difference in the density with which students accompany their school-ratified science reasoning with science discourse.

Our results revealed that at pre-test, students across conditions applied science discourse to 49% of their school-ratified science reasoning instances. There was no significant difference by condition in the density with which students' applied science discourse. At post-test, our results revealed that students significantly increased their rate of science argumentation density across conditions, with 75% of school-ratified science reasoning examples co-occurring with science discourse,  $t(25) = -2.8$ ,  $p = .011$ ). One-way ANOVAs revealed that this increase in density was not significantly different by condition. In other words, across all conditions, students demonstrated a significant increase in their rate of science discourse application from 49% of science

arguments to 75% of science arguments ( $SD = .47$ ).

Taken together, these additional analyses demonstrate that students who heard a science example in AAE were more likely to demonstrate science reasoning at post-test ( $M = 6$ ,  $SD = 2.55$ ) than those who worked with a character that exclusively spoke SAE ( $M = 2.25$ ,  $SD = 1.59$ ). Both groups were equally likely to apply science discourse towards those arguments 75% of the time at post-test, significantly more than at pre-test (49% of the time). Thus, we take these results to mean that students generally demonstrated a stronger density of science argumentation after hearing Jamie's model. However, the students who heard Jamie demonstrate the model in AAE produced *more instances* of these arguments than those who only heard Jamie speak in SAE. We interpret these results to mean that Jamie's dialect may have impacted *how much reasoning* students demonstrated, rather than *what type of discourse* students demonstrated.

#### 4.4.3 Dialect use

Students dialect density was assessed for each of the three recordings students produced: their social introduction, their science pre-test, and their science post-test.

These three variables were averaged together to provide a preliminary understanding of students' dialect across the full experiment. Across all 26 participants, the average DDM was .04 ( $SD = .031$ ). By condition, students who worked with SAE Jamie generally trended to demonstrate a higher dialect density ( $M = .047$ ,  $SD = .04$ ) than those with AAE Jamie ( $M = .024$ ,  $SD = .01$ ). These average distributions are demonstrated in graph [x].

To determine if Jamie's dialect impacted students' dialect, we performed a one-way ANOVA to measure differences in students' DDM in the two task contexts immediately following exposure to Jamie's speech. These tests demonstrated that there was no significant impact of the intervention dialect on students' own dialect production ( $p > .05$ ).

We then evaluated if students' own dialect use moderated the impact of agent dialect. Terry and Smith [Terry, 2010] proposed that if an intervention is resulting in increased cognitive load



for students, there would likely be an interaction effect between the students' own dialect use and their performance within the intervention. A linear model revealed that there was not a significant interaction between students' own dialect use and their condition on their science performance.

#### **4.4.4 Social perception**

Following the intervention, students were given a short survey about their thoughts on their partner. To encourage honest answers, we assured students we wouldn't share their responses with anyone else, and just wanted to make sure we'd be able to give them a partner they really liked next time. Students were asked several questions about Jamie, such as "How fun was your partner?" and "Did you like your partner?" Of these, we were most interested in students' responses to their perceptions of Jamie's intelligence and friendliness. Several other scholars, such as Bernstein [Bernstein, 1960], have found that speakers of stigmatized dialects are more likely to perceive *other* speakers of that dialect as less intelligent than standard speakers, but more friendly. Questions were answered using a 1-4 scale that students were taught how to use prior to taking the survey. Students were also asked several free response questions about Jamie to see if anyone would mention dialect.

Across the board, students almost exclusively selected the maximum option (4) for each of our questions. There was no significant difference in response by condition ( $p > .05$ ). Across students, Jamie's friendliness was rated as an average of 3.7 ( $SD = .89$ ), and her intelligence as 3.8 ( $SD = .33$ ).

On the free response surveys, five of the fifteen students who heard the agent speak AAE at some point (across both the AAE and code-switching conditions) indicated that Jamie did not sound smart. The free response question asked "If you think Jamie sounded smart, why? If not, why not?" We annotated students' responses for whether or not they referred to the character's dialect in some way. See figure [x] for examples of student responses. We performed a chi-square analysis to understand if students who heard the agent speak AAE were significantly more likely

to indicate that Jamie's dialect made her sound *not smart*, but the differences were not significant ( $p > .05$ ).

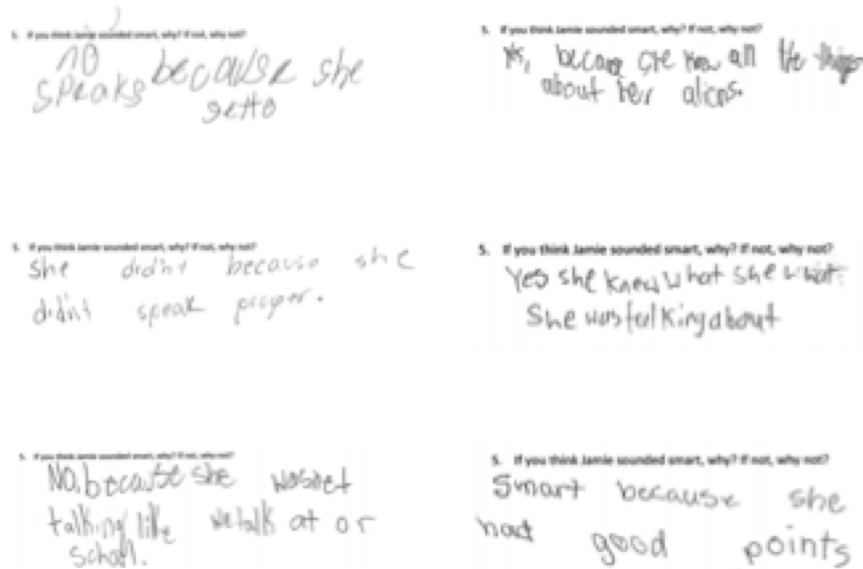


Figure 4.5: Students' perceptions of how smart Jamie sounded when they heard Jamie speak AAE (left) or when Jamie only spoke (SAE). Though Jamie's dialect was only negatively called out in the code-switching and AAE conditions, there was no significant difference of these comments by condition.

## 4.5 Discussion

This study was done to examine the sensitivity of students' science language variables and their dialect use to a short peer example. We also emphasize here that we had only 26 participants, with either 8 and 9 students per condition. Thus, while the results presented here are significant, the small sample size may contribute to a higher potential for Type I (false positive) error.

This study demonstrated that the dialect in which students hear examples of school-ratified science argumentation moderates the level to which students demonstrate their own science rea-

soning and science discourse at post-test. While students across conditions were more likely to communicate their science reasoning within science discourse more after hearing the agent's example (from an average of 49% of the time at pre-test to 75% of the time at post-test), those who heard the model in AAE demonstrated significantly more instances of science reasoning and science discourse (and thus expectedly, science argumentation). However, the small number of average instances of science argumentation among students in the monodialectal SAE condition may limit our ability to interpret that finding. Students in the monodialectal AAE condition demonstrated an average of less than 2 instances of science arguments at post-test. This is a significant increase from their pre-test performance (which averaged at less than one instance), but in practice, there may not be a strong theoretical difference due to such few instances of argument production, even at post-test.

Conversely, students who heard the science example in AAE demonstrated an average of 3.8 (SD = 2.5) instances of science argumentation at post-test. Our interpretation of these results is that agent dialect may have impacted *how much reasoning* students demonstrated at post-test rather than *what type of discourse* students demonstrated. This may mean that rather than *teach students science discourse* or perhaps even *elicit students' use of science discourse at a greater rate*, the impact of agent dialect may have been more specifically impacting how many instances of school-ratified reasoning students demonstrated within their four minute window.

These results also demonstrated that exposure to Jamie's dialect did not have significant immediate impacts on students' own dialects in either the social task or the science task. As a common critique of exposure to dialect in the classroom is that it will increase students' production of dialect [?], this demonstrates that students' dialect production after brief exposures may not be so sensitive. These data also demonstrate that students continue to use some amount of AAE even in an SAE-only environment, adding further to the argument that exposure alone may not result in classroom AAE reduction. Of course, this study was intentionally designed to understand only the *immediate* impacts of *brief* exposure; while these results may indicate

that students' dialect was not sensitive to partner dialect under these conditions, the impact of intervention dialect over a longer span of time remained at this point unclear.

Students who worked with the bidialectal agent heard Jamie's social introduction in AAE, but their science example in AAE. There was no significant difference in science reasoning or discourse among these students at post-test compared to those from either of the monodialectal groups. The range of reasoning gain among students in the SAE-only condition was between "-1" and "4," and the range of reasoning gain among students in the AAE-only condition was between "1" and "8." Conversely, the reasoning gain among those in the code-switching condition ranged from "-3" to "9." This may be due to there being different mechanisms for why an AAE speaking agent helped students, such as some students simply needing the motivation of hearing the agent speak AAE during a social introduction, while others benefited from the dialect in which the science information was provided. It may also just be due to the small number of participants in the sample. This specific question remains an interesting topic for future work.

The social perception results were not significant, perhaps also due to the small number of participants in the sample. However, qualitatively, the presence of 5 students who heard the agent speak AAE volunteering that the agent's dialect contributed to Jamie *not sounding smart* is in line with Kuhne's [Kuhne et al., 2013] finding that dialect-speaking agents are perceived as less intelligent. Similar results have also been found from the social science literature, where standard speakers were generally perceived as intelligent, while dialect speakers were generally perceived as likable [Giles and Marlow, 2011]. However, in this study we have no evidence that students' demonstrated different affective feelings about either version of the agent.

#### **4.5.1 Open questions**

The results demonstrated here indicate that agent dialect does in fact influence the science performance behaviors of African American students. This science task was very unguided; Students were told their assignment and then given four minutes to complete their answers without

prompts for particular responses to give. Students were able to say as much or as little as they wanted within this window. However, differences in students' reasoning performance was not due to students saying *more* in the AAE-only condition. In fact, while these students demonstrated an average increase of 45 words from pre-test to post-test, those with the SAE-only agent demonstrated an average increase of 88 words. Thus, those with the AAE-only agent said less, but a larger percent of what they said was towards the production of science arguments.

Overall, this experiment is in line with the prior literature outlined in the introductory chapters: there appears to be a benefit of bidialectal interventions on students' learning outcomes. What remained to be understood, however, is *why this might be*. Broadly, there have been two competing (though likely synchronous) hypotheses proposed by scholars: *decreased cognitive load* and *increased social alignment*.

Terry et al [?] posited that one way to identify if cognitive comprehension factors are moderating students' performance in a learning environment is to investigate if there is a linear association between students' dialect density and their performance. We did not find any association between students' dialect production and their performance outcomes based on agent dialect condition. While our data does not adduce specific evidence for this cognitive hypothesis, it can not be fully ruled out. It may be the case that our measurement of students' dialect was too crude, as we only measured morphosyntax and not other features of AAE, such as phonological distinctions.

Another possibility is that this analysis failed to capture students' dialect in an appropriate setting. In our analyses, we investigated the potential moderating impact of students' dialect both during a social task (presumably when they would have felt more comfortable producing more dialect) as well as during the science pre-test (which would perhaps be a cleaner measure of the dialect students use in tasks that traditionally require SAE). However, both of these dialect measures were taken after students had already heard the agent speak. It is possible that students demonstrated some amount of accommodation to the agent's dialect (though there was

no significant difference by condition), which muddled this interpretation.

Regarding this study's potential support for social hypotheses, there remain two possible interpretations. Many scholars who posit bidialectal approaches argue that this improves students' social affect and sense of emotional identification to the classroom [Ladson-Billings, 1995b, Sweetland, 2006, Wheeler and Swords, 2004, Mallinson and Child, 2007, Cummins, 1979]. In line with this explanation, it may be that students *liked* the bidialectal agent more, and thus demonstrated more reasoning. This would also be in line with peer emulation theorists [Bandura and Cervone, 1986, Berndt, 1999, Baron and Kenny, 1986], who argue that students are more likely to model peer behaviors produced by a peer who they believe to have *social capital* (e.g., popularity). If students perceived Jamie to be more popular or to have more social capital when he produced science in AAE, they may have been more likely to emulate his model. However, we did not find any difference in students' social assessments of Jamie by condition. This may be because all students marked their perceptions near ceiling. It is possible this social perception questionnaire was not designed in a way that captured any differences in students' perceptions.

An alternative interpretation of this social mechanism is derived from contemporary investigations of stereotype threat theories. According to these theories, some students may respond *better* after being exposed to a potential stereotype that implicates the student. This has been found in more recent analyses of African American students, particularly in situations where there are strong community contexts, such as at a Historically Black University [Solorzano et al., 2000, Steele, 1997]. The authors of these studies interpreted that African American students' increased performance after being exposed to a negative stereotype about African American populations is due to their desire to *disidentify* with the stereotyped community and *defeat* this expectation [Steele, 2003, Aronson et al., 2002]. In other words, this possibility would posit that students who felt *negatively* about Jamie's non-standardized dialect use may have chosen to respond with increased science reasoning as a way to mark their own participation in an academic community (one which Jamie, due to her dialect, was clearly not a part).

In our own results, it is possible there were aspects of this at play. As demonstrated by the free response answers pictured above, a number of the students who heard Jamie speak in AAE wrote that Jamie *doesn't speak correctly* and *wasn't talking like we talk at our [sp: or] school*. Particularly salient in that second example is the use of the word *our*, with that student thus identifying himself as a member of the school community which does not advocate for AAE use [Brown et al., 2017a]. Despite this, each of these three students did demonstrate AAE use themselves throughout the intervention. This is in line with Edwards' [Edwards et al., 2010] proposal that students may not have yet fully acquired the metalinguistic awareness to recognize which dialect features they themselves are producing, though they can recognize them in others' speech.

These results generally support the calls to action provided by proponents of culturally-aligned pedagogy, who argue that integrating students' home dialects into the learning environment may result in benefits such as increased reasoning [Webb and Webb, 2008] or perhaps task engagement [Sweetland, 2006]. They do not however explain *why* students who heard Jamie presented the science model in AAE demonstrated more instances of science discourse, science reasoning, and science argumentation than those who heard the model from an agent who only used Standard English. Gaining additional clarity about these results became the focus of our two following experiments that used a dialogic virtual agent.





## Chapter 5

# Designing a bidialectal, dialogic agent

In the prior four chapters, I have aimed to communicate the arguments that have been made by linguists and cultural theorists who study educational achievement: the language styles explicitly or implicitly included in or validated through a learning environment seem to impact the resulting effectiveness of that learning environment for some students. To my knowledge, these theories about bidialectal pedagogy and the hypothesized benefits for language minority students have not been evaluated with regards to educational technology interventions.

In this chapter, I first provide a brief overview of other educational technologies that have integrated aspects of students' culture into their design. While some of these systems have acknowledged and manipulated non-standard dialect features as part of the design, the language style used was typically one of several features manipulated. While the complex and integrated facets of culture mean that this was likely a *positive* choice from an intervention design standpoint, it limits our ability to identify the relative benefits of various specific choices. Regardless, to better situate Alex in a body of work investigating *culturally sustaining educational technologies*, I begin with a brief overview of the state of this field.

Following that, I then provide an overview of virtual peer system and our experimental design paradigms. I describe the science activities we designed into the agent based off of our iterative pilot testing, and present an overview of what prototypical exchanges look like with the agent.

Finally, I conclude by scoping out the specific research questions we address in this work. I also outline the interactional factors I hypothesized might *mediate* the relationship between agent dialect and student performance, including several metrics of task participation [Wheeler and Swords, 2004, Fogel and Ehri, 2000, Sweetland, 2008] and social behavior [Boykin, 1994, Fordham, 1999, Brown et al., 2017a]. I also outline the student factors I hypothesized might *moderate* the impact of agent dialect, such as students' own dialect use [Terry et al., 2010, ?], students' dialect attitudes [Ogbu, 1991, ?, Brown, 2006], and students' standardized reading level [Craig et al., 2009a, Godley and Minnici, 2008, Charity et al., 2004a].

## 5.1 Alex: a bidialectal virtual agent

In this chapter, I overview the design of our virtual peer, the design of our science activities, and the design of our experimental paradigm.



Figure 5.1: Alex, our virtual peer experimentation platform, talking with a 2nd grade student

### 5.1.1 Experimental paradigm

Alex was set up in an isolated location in our partner schools (e.g., a small conference room or an empty gymnasium), and an experimenter pulled students out of class one at a time to work with the agent independently for about 20 minutes each session. Science activities were designed to be open-ended discussions that were in line with the broad science standards for students in 2nd and 3rd grade. We designed the science activities to involve two types of dialogue tasks within each session (10 minutes each). For each student, the intervention involved both an informal peer task that involved brainstorming ideas together, followed by a more formal teacher task that involved taking turns asking each other questions in preparation for their subsequent final presentation. In the monodialectal condition, Alex spoke Standard American English (SAE) during both tasks. In the bidialectal condition, Alex spoke African American English (AAE) during the peer brainstorming task, and then switched to SAE during the teacher presentation practice task.

This intervention choice of first brainstorming and then reviewing answers was done for two reasons. First, it allowed a manipulation of *formality* which would provide a reason for Alex to demonstrate bidialectalism. As prior research evidenced that African American college students felt positively about recordings of characters that demonstrated code switching according to broad social norms but *negatively* about characters that demonstrated code switching *away* from social norms [Koch et al., 2001], the bidialectal character demonstrated AAE during the peer brainstorming task, and then switched to SAE during the teacher presentation task. Second, we also made this design choice because it allowed Alex to speak with identical dialects during the teacher presentation task in both conditions, which allowed for a cleaner comparison across conditions.

As described in Chapter 3, we designed the science activities with the virtual agent to involve two tasks that varied in formality as a way of demonstrating bidialectalism in context. For each student, the intervention involved both an informal peer task that involved brainstorming ideas

together, and a more formal teacher task that involved taking turns asking each other questions in preparation for their subsequent final presentation. In our instantiation of the *business as usual* virtual agent, Alex demonstrated monodialectal use of SAE in both task contexts. In our instantiation of the *AAE-integrated* virtual agent, Alex bidialectally demonstrated AAE in the informal peer task and SAE in the formal teacher task. In doing so, the AAE-integrated agent demonstrates a form of *code-switching*, where the formality of the task context influenced the agent’s dialect production. The peer task took place over the first ten minutes of the intervention, and the teacher task took place over the remaining ten minutes of the intervention. Thus, to clarify, for the first ten minutes of the intervention, students in the monodialectal condition heard Alex speak SAE and those in the bidialectal condition heard Alex speak AAE. However, during the teacher task over the remaining half of the intervention, Alex used SAE in both conditions. In other words, Alex being identical during the last ten minutes of the intervention across both conditions.

### **5.1.2 Guided science activities**

We designed three science tasks that Alex is able to complete. For each one, we first investigated real students participating in the task to build a corpus of what realistic child-child dialogue looks like within the domain and underwent several iterations of piloting our utterance library. For each utterance we added into the agent’s dialogue bank, the process included recording it in both AAE and SAE with our bidialectal voice actress, generating lip syncing behaviors for both dialect versions so that the language looked realistically spoken by the agent, and then modeling nonverbal behaviors that accompanied the language. Thus, the process of building each task was quite lengthy, and often required several iterations of recording new utterances based on unexpected contributions presented by students in our piloting sessions. The three tasks we completed for Alex to participate in with students are overviewed below.

The Creature Task depicted a fictional creature within its natural environment. Students were

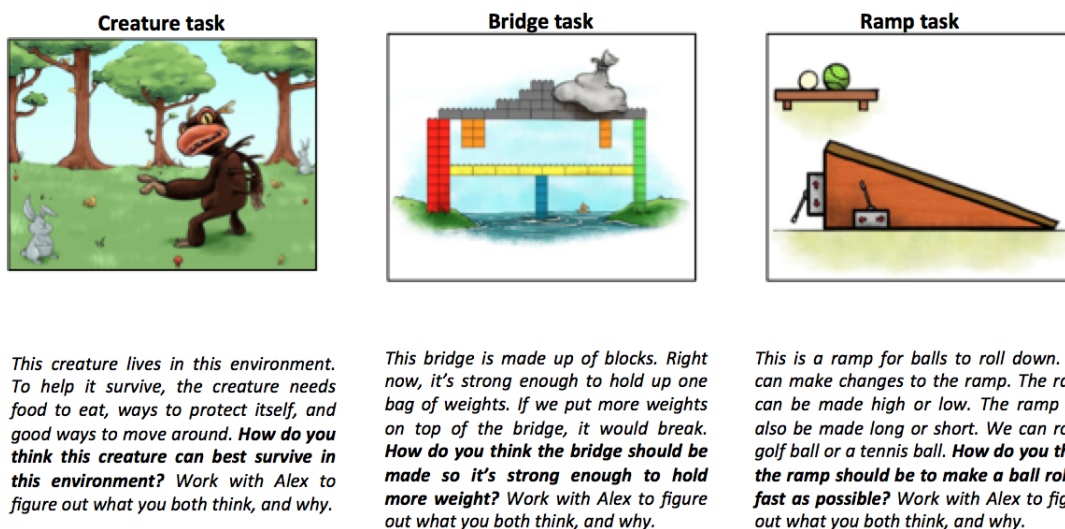


Figure 5.2: The three science tasks Alex works on with students

asked to make inferences about the creature's survival habits with regard to food, protection, and movement based upon its appearance and habitat pictured. Alex guided questions in this task like *"How do you think the creature can protect himself in the night time?"* and *"Where do you think the creature sleeps?"* Alex focused on building dialogue with students about how aspects of the creature's physical attributes could be applied to promote survival (e.g., *"I think he probably only eats worms because he has long nails to help him dig holes, and his fingers are long enough to help him pick up things really easily so he doesn't have to fight."*)

The Bridge Task shows a girder and beam bridge made from square blocks. The abutments and piers are of unequal widths and are spaced unequally. Students are asked about structural changes that would allow the bridge to support more weight. Alex guided questions in this task like *"Which beams should we make thicker to help it hold up the weight in the middle better?"* and *"Do you think we should connect this middle beam to anything?"* Alex focused on building dialogue with students about how design choices regarding symmetry, bottom-heavy building choices, and low centers of gravity could maintain stability (e.g., *"I think we should probably make these side beams lower so that there's less wobbling room at the top when we put new*

weights on it.”)

The Ramp Task depicted a ramp with a tennis or golf ball that could roll down the ramp. The length and degree of incline of the ramp are shown to be adjustable. The students are asked to determine how the ramp parameters could be set to maximize for speed of a ball rolling down the ramp. Alex guided questions in this task like “*Do you think the ball goes faster when the ramp starts high up or low?*” and “*Do you think the fuzzy stuff on the tennis ball would make it go faster or slower?*” Alex focused on building dialogue with students about how longer and higher ramp designs would promote ball speed, especially with the heavier ball (e.g., “I think that when the ball is heavy, it’s gonna get faster and faster with every roll because then the gravity can keep pulling it down, so we probably should pick a long ramp and the heavy ball to make it fastest.”)

### 5.1.3 Dialogue design

The agent’s language was designed to demonstrate the principles of science discourse as defined by the NGSS. As such, utterances by the agent demonstrate science vocabulary and reasoning structures. However, while they were chosen to model these features, all instances of speech were taken directly from actual utterances spoken by students performing these tasks during pilot studies and then adapted to fit our operationalizations of science discourse defined more below.

Alex’s dialogue was pre-recorded by a professional African American bidialectal voice actress named Siovhan Christensen.<sup>1</sup> During our recording sessions, Siovhan worked alongside a linguistics post-doc to help ensure that the dialectal differences between the AAE and SAE utterances in the agent were consistent, sounded authentic, and followed grammatical structures that would sound unmarked to a native speaker. This collaboration was instrumental to our utterance design process as I am not a native speaker of AAE. Frequently, several times within a single

<sup>1</sup>Siovhan Christensen (not a pseudonym) worked as our voice talent, and she has appeared in both film and television productions. Her work can be viewed here: <http://vergiasiovhan.wixsite.com/mysite>

recording session Siovhana would identify that an AAE utterance we had written based on our understanding of dialect rules just *didn't sound right*, and in these instances, we would follow her recommendations for restructuring the utterance to sound realistic.

After our science tasks were completed after multiple rounds of iteration, Alex was able to demonstrate 1,020 utterances. The break down of Alex's utterances are shown in the figure below.

Category	Category	Sub-category	Total	Examples
Guiding (355)	Collaboration (163)	Co-construction	48	<i>That's what I was thinking too! Because then...</i>
		Reasoning prompts	60	<i>Wait, why's that?</i>
		Follow-up	55	<i>That makes sense</i>
	Alignment (192)	Back-channel	39	<i>Uh huh?</i>
		General	115	<i>The first one</i>
		Clarification	38	<i>Hmm... so which one?</i>
Domain (416)	Creature (153)	Contribution	79	<i>I think he's probably a good climber because he has long arms and short legs, like a monkey</i>
		Elicitation	74	<i>How do you think he protects himself from those like bigger, spikier predators?</i>
	Bridge (132)	Contribution	87	<i>I think that if we keep making it taller, we need to add another supporting beam so it stays stable</i>
		Elicitation	45	<i>Wait, do they even need to be the same height?</i>
	Ramp (131)	Contribution	71	<i>I think when it's longer, then the ball will get more rolling room, and that can make it faster</i>
		Elicitation	60	<i>I wonder if it changes based on how long the ramp is</i>
Social (249)	Scripted (100)	Self disclosures	65	<i>I like playing video games and stuff like that</i>
		Elicitations	35	<i>What were you doing in class before you came here?</i>
	Unscripted (149)	General	101	<i>I did once last year I think. What about you?</i>
		Deflection	48	<i>It's different when you live in a computer world like me</i>

Figure 5.3: Category counts of Alex's dialogue options

As shown in figure 4.x, in addition to domain task utterances, a significant portion of Alex's utterances served to provide social fluidity. While there were several scripted exchanges Alex had with students to build a social connection in small ways (e.g., asking all students in the very beginning of the interaction what they had just been doing in class), many social exchanges were introduced by students. Through our pilot testing, we identified that Alex's failure to comply appropriately to some of these requests could derail the dialogue in ways that were difficult to

recover. For this reason, we added 149 utterances into the system that aimed to be appropriate blanket responses to students' questions either through being broadly general enough to seem like an appropriate reply (e.g., "I did once last year"), or to deflect the question (e.g., "It's different when you live in a computer world like me.") As I describe more in the social measures section below, these responses (and how well students accepted them as answers) seemed to play a strong role in the effectiveness of the intervention for some students.

Alex was operated using a *Wizard of Oz* approach where an experimenter, hidden from the students' view and listening in through a web cam and microphone, would choose Alex's response out of the available set. In the majority of our piloting and each of the experiments presented here, I served as the controller of Alex's dialogue options. To promote consistency across conditions, I had designed a dialogue tree to plan out how Alex would generally respond to different student utterances at various stages of the interaction. However, because of the intention for dialogue with Alex to seem fluid and peer like, and due to the complex nature of free speech, I aimed to make dialogue choices that prioritized communicative fluency rather than using the dialogue tree guide as a prescriptive pattern.

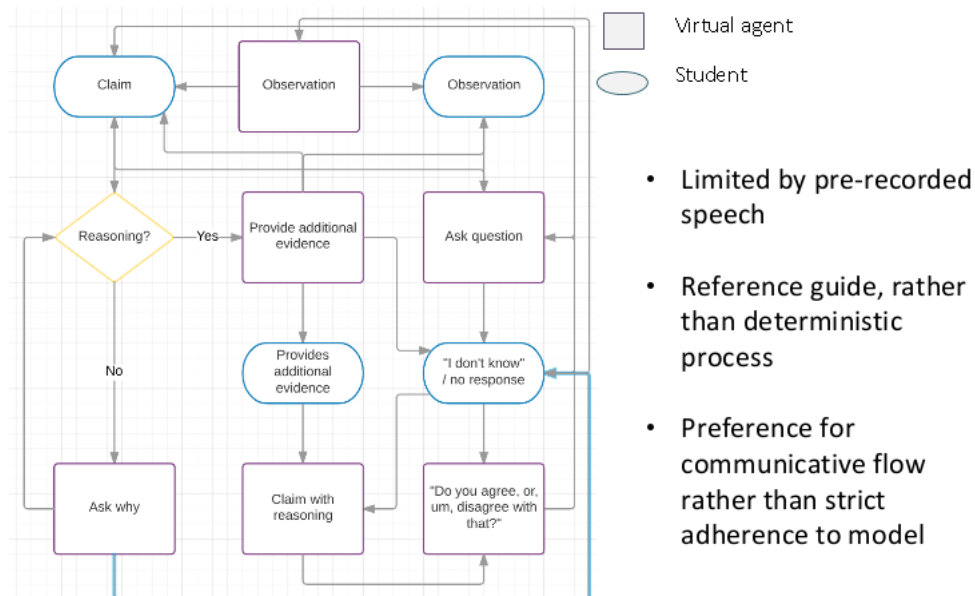


Figure 5.4: General scheme for dialogue between the agent and student



The benefits of this methodology were that due to our frequent piloting during the task design stage and my personal extended practice operating Alex during these dialogues, many of the exchanges between students and Alex were surprisingly fluent. However, it introduces two limitations. The first is that as I operated as both the dialogue operator as well as the primary investigator of this work, there were opportunities for bias in my response choices. After each experiment, we performed post-hoc tests to confirm that there were equal numbers of utterance types produced by Alex across conditions. As I will present in each experiment described below, there were no differences by condition in terms of how many utterance types Alex demonstrated in any of our post-hoc analyses. However, these dialogue options *were* correlated with rapport, regardless of condition. I describe why this might be, and the potential implications on our results, in the discussion sections of the following experiments.

The second limitation of our agent dialogue design was also a result of the pre-recorded nature of Alex’s responses, though this is true of both conditions. While the scripted responses allow the agent to respond appropriately and maintain fluent conversation when on-task, the responses do not always provide for direct response to off-task discussion and questioning. Though we aimed to account for this through piloting, there were occasions where students made conversational demands of Alex that he was unable to answer. We describe this further in our measurements section below.

#### **5.1.4 Prototypical agent dialogues**

To help provide context, I present here prototypical examples of what unmarked (or “expected”) interactions looked like between students and Alex. As I describe in the next session, there were variations in these interactions based on what may have been differences in students’ perceptions, differences in students’ interactional goals, differences in students’ personalities, or other factors we were not able to quantitatively confirm. However, to better explain these variations, we first present typical exchanges during various stages of the agent interaction. Each of these

excerpts are taken directly from real student interactions that are representative for the majority of unmarked interactions with the agent. These examples are taken from the bidialectal agent to demonstrate the use of Alex's dialect, however I demonstrate in bold what the utterance would have been in the SAE condition.

### Initial task introduction

*Alex:* Hi. How's is going?

*Student:* Hi? Good?

*Alex:* Me too. I'm Alex. What's your name?

*Student:* Hi Alex. Um, Destiny.

*Alex:* Hi. What **[were / was]** you doing in class before you got here?

*Student:* Um, reading?

*Alex:* Cool. The rest of my class is in **[their / they]** math groups now.

*Student:* Oh. Do you like math?

*Alex:* It's okay. My teacher **[pulled / pull]** me out to do this. You?

*Student:* Oh, yeah, the lady pulled me out too.

*Alex:* I think **[we're / we]** supposed to figure out how the creature like, **[eats / eat]** and **[moves / move]** around and stuff.

*Student:* Um, yeah I think so too.

*Alex:* Okay... the **[spikes / spike]** on his back, they look almost like a porcupine, but not as sharp.

### Brainstorming in the peer task

*Alex:* What do you think those **[spikes are / spoke]** for?

*Student:* Protecting hisself?

*Alex:* Me too, 'cause then if an animal [**comes / come**] after him, he can stop them from running into him

*Student:* Yeah, and then the bigger animals ain't getting him either

*Alex:* Do you think he can protect [**himself / hisself**] from the bigger animals, um... predators?

*Student:* Yeah I think he can 'cause that's what he use the spikes for, the predators, and staying safe from them

*Alex:* [**Do / -**] you think he [**has / got**] other ways to stay safe?

*Student:* I think um - night probably?

*Alex:* [**Why's / why**] that?

*Student:* I think he go out at night 'cause he got that stripe in his eye for seeing at night, like for night vision

*Alex:* Yeah! And because [**his fur is / he fur**] dark, then [**- / ain't**] no one can see him!

*Student:* Me too! And then he can uh camouflage

*Alex:* That makes sense. I think he can camouflage at night because he [**has / got**] that stripe in his eye, like a cat... and with cats, I think the stripe [**helps / help**] them see at night

### **Transitioning to the teacher task**

*Alex:* I think [**we're / we**] doing good.

*Student:* Me too! I like the ideas.

*Alex:* I think now we gotta practice our final presentations for the teacher.

*Student:* Okay.

*Alex:* Wanna take turns asking each other these questions? First I'll pretend to be the

teacher and I'll ask you questions, and then you can be the teacher and you can ask me questions!

*Student:* Like the questions we've been asking?

*Alex:* Yeah. Let's ask them like a teacher would ask them.

*Student:* Okay.

*Alex:* My teacher likes it when I use my school English whenever I have to do presentation stuff like this.

*Student:* Yeah.

*Alex:* Yeah. Okay, first question...

### **Teacher presentation practice task**

*Alex:* What sorts of food do you think the creature eats, and why?

*Student:* Uh, I think... worms? And the rabbits.

*Alex:* Uh huh? And why?

*Student:* I think he can eat the rabbits 'cause - because - no he can't eat the bunnies because he ain't got legs fast enough to ca- to catch them.

*Alex:* Good answer! I think so too. I don't think the creature looks like he can run very fast because his legs are so short.

*Student:* Yeah. My question is, why do you think he - what can he do with those short legs?

*Alex:* Hmmm... I think the creature looks like a very good climber, because he has long arms and short legs like a monkey, and because he has good toes for gripping on stuff.

Alex followed this script outline for students in both the code-switching and monodialectal conditions. For the monodialectal agent, there was no change to the character's dialect. For

the code-switching agent, the character stopped demonstrating AAE features at this point in the dialogue. From this point forward, for the last ten minutes of the interaction, Alex was identical in both conditions.

## 5.2 Rapport

Krämer [2008] has demonstrated that agent dialect impacts its social reception. Some scholars additionally argue that part of the impact of culturally-aligned learning environments is due to the social importance of feeling your language style is respected and represented Sweetland [2006]. Boykin and Noguera [2011] even more specifically contends that collaboration among peers that seem *similar* is especially important for the academic performance of African American children. If part of the success of linguistically-aligned interventions is due to social factors such as these, we would expect differences in students’ social behaviors toward the agent to emerge over the course of a collaborative dialogue, and for these social differences to be associated with students’ performance differences.

Rapport was evaluated using a “thin-slice” approach, with rapport measured on a Likert scale from 1-7, from the lowest possible rapport to the highest possible rapport. Rather than measure specific behavioral cues as markers of rapport, these measurements provided a holistic quantitation of rapport. Thin slice approaches are based upon the principle that minimal amounts of data can lead to rapid and accurate inferences in judging the overall state of a situation.

Videos of the interactions were divided in 30-second slices and provided to undergraduate interns who served as third-party raters. The video slices were also shuffled in order, such that the raters were not biased by the change in rapport from the previous slice. Three raters were able to achieve inter-rater reliability of Krippendorff’s  $\alpha = 0.89$  for each experiment discussed within the respective section for each experiment), with group discussion to resolve any discrepant ratings. Once inter-rater reliability was achieved, all slices were each rated by a single rater, including re-annotation of the original subset used for training.

We compiled the thin slice ratings into an *average rapport* score during both the peer task and the teacher task for each student. We additionally calculated utopy over the course of the full session to assess the change over time. Utopy measures the likelihood that a dynamic variable such as rapport is increasing from one thin slice to the next. This method was introduced by Sinha et al (2017 cite). Scores closer to 0 indicate that rapport is not generally increasing over time; scores closer to 1 or higher indicate that rapport is more likely to be increasing over time.

On a scale of 1-7, '4' was anchored to be *expected* levels of positive affect among students working on a science task with a peer. The following three example frames are from three 30-second thin-slice clips where rapport was annotated as 1, 4, and 7 respectively.



Figure 5.5: Frames demonstrating rapport at values 1 (low), 4 (neutral), and 7 (high)

### 5.3 Dialect attitudes

Earlier work has found that high school students have varied ideologies about what it means to speak AAE [Godley and Minnici, 2008, ?]. Ogbu describes students' assertions that speaking SAE in certain situations would mark them as *other* among their peers, and that the decision to

use AAE was a marker of community affiliation. Godley's work adds additional nuance, uncovering that traditionally successful African American high school students remark that AAE is inappropriate, and that they do not identify as speakers of this dialect. Sweetland [2006] added upon these results by talking to elementary aged students about their ideologies regarding dialect after having participated in a language variation curriculum, and many identified as successful code-switchers who spoke both AAE and SAE based on the context after the intervention took place. Eckert [2000] proposes that early adolescence marks the time when students begin prioritizing their peer community language expectations over the traditional language expectations of the classroom. For this reason, additional work understanding the ideologies of elementary aged children (especially prior to receiving a language variation curriculum) might provide interesting insights the potential social factors that may be guiding language use decisions or behaviors among these students.

We wanted to investigate students' language attitudes towards SAE and AAE before and after their interaction with our virtual agents to evaluate if language exposure impacted their reported ideologies. Unlike Sweetland's intervention which was an explicit language variation curriculum, this analysis would reveal if simply exposing students to a technology that demonstrated language variation (even without any direct instruction about how to code-switch) could have an impact on student attitudes.

We were sensitive to the concern that by elementary school, students would know that their teachers expected them to use SAE. For this reason, we were concerned with the possibilities for students' experimenter expectation bias. To try and counteract this potentiality, we built a computer application that asked students about their language ideologies without an adult in the room during both pre-test and post-test. The application was voiced by a ten year old African American bidialectal voice actress to promote feelings of *peerness* and social safety within the application. Prior to working with the application independently, the program walked students through an example of how to fill out Likert scales. The experimenter was in the room during

this training so that the student could ask questions if they were unsure of what to do. After the student completed the training, the experimenter told the student that they could fill out this next part on their own, but that they would be outside if they had any questions. Once the student confirmed that they understood, the experimenter left the room.

The ideologies application then gave the students the following instructions. *“Now I wanna ask you some questions about the way people talk! Lots of kids speak in lots of different ways. One way is called Standard. Another way is called Dialect. Some kids only speak one or the other. Some kids speak both. If I were speaking Standard I would say ‘my brother is always telling me he doesn’t have any candy, but I know he does!’ If I was speaking Dialect I would say ‘my brother always be telling me he don’t got no candy, but I know he do!’ Hear the difference? These next questions are about speaking Standard and speaking Dialect.”*

The application then showed the students 13 questions about using SAE or AAE in different settings, referring to them as Standard and Dialect. The questions were of the form *“sometimes it’s good for kids to speak **Dialect** in the classroom”* or *“sometimes it’s good for kids to speak **Standard** when they’re hanging out with their friends.”* Each question had a Likert scale option from 1 - 5 that was identical to the one shown during their training, where 1 was labeled *“definitely no!”* and 5 was labeled *“definitely yes!”* At any time, the students could click a *help!* button that would re-play the examples of SAE and AAE sentences provided in the instructions.

Due to the inherent flexibility and variation of language, we began all questions with *sometimes* to intentionally measure students’ ideologies about *whether or not this dialect was ever able to be an option in this setting*. Thus, a score of “5” would mean *“I believe that there are **some** situations where this dialect is okay in this setting.”* A score of “1” would mean *“this dialect is **never** okay in this setting.”* Because of this built-in hedging, we identified that students with *positive language ideologies* would mark “5” for both AAE and SAE. In other words, a score of “5” would simply mean *“there are some situations in which this is okay,”* thus making it reasonable for students to have “5” for both questions about SAE and AAE.



The following questions within each dialect category were averaged together to produce a score for students' language ideologies within that category.

(1/2) Sometimes it's good for kids to use [Dialect/ Standard] when they're just hanging out.

(3/4) Sometimes it's good for kids to use [Dialect/ Standard] when they're in the classroom.

(5/6) Sometimes kids sound smart when they speak [Dialect/ Standard].

(7/8) Sometimes it's good for kids to know how to speak [Dialect/ Standard].

Two questions were included to measure students' expectations about dialect in media that did not contribute to ideology scores. These were:

(9) Sometimes it's good for books to use Dialect

(10) Sometimes it's good for computers to use Dialect

Two questions were used to specifically measure students' ideologies about bidialectalism, including:

(11) It's good for students to know how to speak both Standard *and* Dialect.

(12) It's good for students to use both Standard *and* Dialect in the classroom.

Finally, the last question on this survey asked students "I use dialect sometimes." This was not included in language ideology measures, but to assess students' self-identification as a dialect speaker.

To ensure our instructions were sufficient for 2nd - 4th grade students to understand what we meant by *Dialect* and *Standard*, we piloted this application as part of our virtual agent piloting. We held in-lab family / friend focus groups where we spoke to sixteen elementary aged students across five different groups. We chose to include parents such that if students *didn't* know what we meant by Standard and Dialect, we could see how parents then explained it to their students. This would also ensure that we had parent buy-in for the language we were using for this assessment, and that if not, we could learn from their suggestions about how they would recommend educators talking with their students about this kind of language variation.

After playing students the instructions, we asked them if they had ever heard the words *dialect*

and *standard* before, and made the space for them to choose to respond to the instructions. We did not ask about them directly at this stage. Then, we played voice clips from our virtual agent in either SAE or AAE. We told the students that Alex could speak both ways, and asked which way they thought Alex was speaking. While students did not always correctly identify Alex's sentences correctly on an individual instance level, students largely identified Alex as speaking Dialect when he used a salient AAE feature. Regardless, there were no instances where one of our pilot students did not demonstrate awareness that they knew what the instructions meant *categorically* by Dialect and Standard.

## Chapter 6

### Study 2: Interaction and transfer

Our prior study demonstrated that students who heard a science presentation from a monodialectal African American English speaker demonstrated more instances of science discourse at post-test than students who heard the presentation from a monodialectal Standard American English speaker. However, these presentations were given immediately after students heard the agent's model; for this reason, it is possible the effect we found might not have been *learning* in a deep sense, but rather the result of sociolinguistic elicitation or subconscious accommodation. Similarly, because of the non-interactive nature of the design, it was difficult to gain insights into what mechanisms may have been driving that relationship.

First, we investigated students' use of science discourse during a pre-test the day before the intervention, and a post-test the day after the intervention. We posited that any differences on a next-day post-test would better represent *learning* than our prior study. *Would agent dialect impact students' use of science discourse the next day?*

Second, we investigated students' rapport with Alex during the intervention. As reviewed in the methods chapter, scholars who promote the use of culturally aligned pedagogies often list the social benefits of these interventions, such as through students' increased feelings of representation. We posited that students would demonstrate increased positive outward affect (here simply called *rapport*) with a bidialectal agent.

Our final primary question under investigation was whether rapport would mediate all or part of any relationship observed between condition and post-test science performance. Social factors have been argued to impact the effectiveness of a peer learning environment, such as with students demonstrating “non-adaptive behaviors” when working with partners they didn’t care for. Similarly, the pedagogical agent design has been argued to impact students’ social perceptions, and in turn, the level to which they benefited from the intervention. We hypothesized that students who demonstrated more rapport with Alex would also demonstrate more science discourse with Alex, thus improving the effectiveness of the intervention.

Further, we hypothesized that rapport may also impact students’ post-test science discourse directly, not just through its potential for increasing practice with the agent. I posited that such a finding would be in line with what Carol Lee refers to as a *cultural apprenticeship* effect, where students directly and personally benefit from the experience of working on an academic task productively with peers who they identify with [Lee et al., 2005, Brown et al., 2017b].

## 6.1 Study overview

We partnered with a local, no-cost charter school that primarily serves African American students. The school primarily recruits students from under-served neighborhoods, and reports that 99% of students receive free or reduced lunch. Their website expresses that their science mission is to design lessons which create “rich, academically oriented oral language experiences that promote receptive and expressive language skills.”

Consent forms were distributed to 2nd and 3rd grade students. Due to the low return from 3rd graders, we chose to work exclusively with the 2nd graders. Of the 42 students enrolled in 2nd grade, we received signed consent forms from 30 students. All students enrolled in our study were African American. All students demonstrated some use of grammatical features indicative of both SAE and AAE in at least one sample of recorded speech during this experiment.

Of the 30 consented participants, there were 15 boys and 15 girls. Participants were randomly

assigned to condition controlling for gender. Students had taken their standardized reading assessments the week prior to the study, and we did not receive their scores until the rest of our data had already been collected and annotated. Retrospectively, 8 girls were reading at or above reading level, and 7 girls were reading one or two grades below reading level. Among the boys, 4 were reading at or above reading level, and 11 were reading one to two grades below reading level. Within those who had been assigned to the monodialectal

Each session with Alex involved a peer brainstorming task (the first ten minutes) and a teacher presentation practice task (the remaining ten minutes). The goal of the peer brainstorming portion of the intervention was to demonstrate arguments for and against five open-questions in the task. In this single-session study, Alex worked with each student on *the creature task*, an open-ended science activity designed to promote argumentation about how an imaginary creature could best demonstrate adaptation to its environment. Though the task was intentionally open-ended to promote natural peer-like dialogue, Alex aimed to guide the dialogue to address five science questions within the domain, such as what the creature might eat (bunnies or worms) or the creature's sleeping patterns (nocturnal or diurnal). Each question had two most likely answers based on the design of the environment, and Alex was able to demonstrate arguments for or against each of them.

Students' participation in this study took place over three days. On days one and three, students were given pre and post assessments of their science performance, dialect use, and language ideologies. These assessments were given by a white experimenter who identified herself to the students as a local teacher who was testing how students do science. On the second day, students were pulled out of class to interact with Alex for a single 20 minute session. The experimenter who pulled students out of class was African American, and did not mention any association with the confederate teacher the students saw the day before. She was blind to condition, and did not know which version of Alex students were working with. She explained to the students that "*this was Alex*," and that the two of them would be talking together to come

up with their hypotheses about how this creature interacts in this environment. This task was similar to the pre-test students completed the day before.

### **6.1.1 Research questions**

In this study, we build off of these results through an experimental paradigm using Alex.

Our primary two research questions are address throughout this work are:

1. Do students improve their use of science discourse at post-test?
2. Does Alex's dialect moderate how *much* students improve their use of science discourse at post-test?

We additionally investigate how Alex's dialect might impact students during the intervention. We hypothesized that by condition, we would see differences in student behavior in two ways.

1. Students with the bidialectal agent would demonstrate more science discourse with Alex
2. Students with the bidialectal agent would demonstrate higher rapport with Alex.

### **6.1.2 Science discourse assessment**

To assess students' science reasoning and science discourse at pre- and post-test, students were given a science task by the experimenter. The task was designed to measure the same type of underlying skill that Alex aimed to scaffold in the intervention: production of hypotheses supported by evidence from a task. Like with the task Alex spoke about with students, the content of the task was based around the idea that animals can use their body to help them survive in an environment. This is in line with the Next Generation Science Standards for 2nd and 3rd grade [The Next Generation Science Standards For States, 2014]. As we wanted to measure transfer rather than memorization, we aimed to make sure that the creatures and environments given in the pre- and post-tasks were distinct from the ones present in the intervention. This prevents students from simply reiterating information that was said in the intervention on the post-test.

The results from the previous dialect elicitation study found that students who heard the science task in AAE demonstrated *more instances* of reasoning during a free-form monologue task. In this study, we chose an assessment design that would provide some additional control on the *amount* of reasoning students were given the opportunity to produce. Thus, we could better understand students' use of science discourse during five specific questions, rather than in a more open ended monologue environment. In doing so, we aimed to see if students' improved the percentage of their responses which used science discourse.

The pre-tests and post-tests were provided to students by the experimenter who had introduced herself as *a teacher from Pittsburgh Public Schools who was here to see how students do science*. We wanted to ensure that the task *felt like a traditional instantiation of school*, and so the experimenter was instructed to maintain professionalism through her interaction with the students.

The instructions for this task were as follows: *This is an imaginary creature who needs to live in one of these environments. For each environment, I'm going to ask if you think it would be a good or a bad place for this creature to live in, and why. Then, at the end, I'll ask which is the very best environment for the creature, and why. Okay?* The experimenter then asked the students, for each environment, *"Would this be a good or a bad environment for this creature to live in, and why?"* The experimenter was instructed to say *"Uh huh?"* at each student pause to offer the student the opportunity to continue speaking if they chose. If they didn't respond to that prompt, the experimenter would ask *"Anything else you want to say about this environment?"* When the student was finished, the experimenter would ask about the next environment. At the end, the experimenter said *"Out of all of these environments, which one would be the best for this creature to live in, and why?"* The experimentation script instructed the experimenter to stop the student and go to the next question if any individual task response went on for longer than two minutes, but that did not happen during either the pre or post-test. The experiment was blind to condition during both the pre- and post-test.

Across each of the five questions at pre- and post-test, we annotated students' responses for instances of science reasoning and science discourse. We used these measures to investigate two science performance outcomes: (1) the total number of science discourse instances demonstrated, and (2) the density of science discourse applied to instances of science reasoning. We refer to these two measures as "science discourse" and "science argument density" respectively.



Figure 6.1: Pre- and post- science tasks to measure reasoning and discourse

### 6.1.3 Baseline dialect assessment

Students were given an assessment to measure their “baseline” dialect during the pre-test, during a task designed to elicit students’ production of comfortable speech. We emphasize here that dialect use is a complex process dependent on many factors, such as interlocutor, context, task genre, and so forth [Craig and Washington, 2004]. Craig and Washington [Craig and Washington, 2004] assessed which of several contexts resulted in elementary-aged students demonstrating the *most* AAE features. They found that students generally used the most AAE when describing a provided picture with some complex interactions, taken from Bracken’s expressive language task [Bracken, 1984]. They proposed that this might be due to the increased grammatical complexity of describing interactions, as well as the potential for reduced attention to language that might be occurring during tasks that seem more aligned with school, or in dialogue with an experimenter. Charity [Charity et al., 2004a] also identified that students generally demonstrate more AAE dur-



ing longer, more complex sentences than shorter and simpler ones. For this reason, students were asked to complete a picture description task as their first assignment at pre-test. Students were instructed to describe what they saw going on in the picture, and could talk about anything they thought might be happening. Students were told they had four minutes to talk about the picture, and that they could keep describing everything they saw in the picture until the experimenter returned. The experimenter left the room and returned four minutes later to stop the recording.

Analyses by Craig and Washington [Craig and Washington, 2004] and Charity [Charity et al., 2004b] both indicated that it isn't students' *use of dialect in any context* that is predictive of reduced academic performance, but only the use of dialect in specific contexts that called for SAE. In fact, Craig and Washington [Craig and Washington, 2004] demonstrated that students' demonstration of task-based code-switching was *positively* predictive of academic performance. Thus, the two primary dialect use variables we investigate in our analyses below are students' use of AAE during the science pre-test (when, presumably, students might expect the situation to call for SAE), and their demonstration of code-switching between the picture description task and the science pre-test. The code-switching variable represented the delta in students' dialect production between both contexts, with higher variables indicating more task-based code-switching.

We measured students' dialect by calculating a density dialect measure (DDM) during each task. This annotation procedure is described more thoroughly in Chapter 3's methods section, but we additionally summarize it here. The DDM measure was calculated on eight grammatical constructions which are frequent markers of AAE, and which we observed frequently among the students in our target demographic through our qualitative studies and previous dialect elicitation study. For each of these grammatical constructs, we annotated the students' production of either the *SAE* or *AAE* production of this feature, and measured the percentage of opportunities to use this feature in which the student chose the AAE usage. Because there is substantial overlap between AAE and SAE, and complex linguistic and social reasons why one might choose one grammatical construction over another, it's important to remember that DDM calculations repre-

sent a simplified calculation of students' dialectal production. We calculated a DDM during each instance of students' language production, including their pre- and post-test science assessments and their interactions with Alex during the peer and teacher tasks.

## 6.2 Overview

For clarity, we re-state here that students' 20 minute dialogues with Alex all involved a ten minute informal peer task and a ten minute formal teacher presentation practice task. Students with the monodialectal SAE agent heard Alex demonstrate SAE in both the peer and teacher tasks. Students with the bidialectal code-switching agent heard Alex demonstrate AAE in the peer task and SAE in the teacher task. We emphasize that Alex's dialect was thus identical across both conditions during the teacher presentation practice dialogue over the second half of the session.

I report here an overview of students' pre-test scores to better illustrate our student population. Due to equipment failure, we lost the audio for the picture description task and science pre-test for three participants at pre-test. In each of these three cases, students' attitude surveys, which were taken through a different app on that same computer, were recoverable. Unfortunately, each of these three students had been randomly assigned to the bidialectal condition. I do not believe there is a systematic reason behind this coincidence.

Student pre-test factors by condition								
Condition		Developmental Reading Assessment score	Dialect (picture description)	Dialect (science pre-test)	Code-switching at pre-test	Ideologies about Standard English (pre-test)	Ideologies about Dialect (pre-test)	Reasoning with Science Discourse (pre-test)
Monodialectal agent	Mean	22.67	.098	.092	.006	3.85	3.28	.42
	N	15	15	15	15	15	15	15
	Std. Deviation	7.51	.10	.08	.13	.68	.82	.30
Bidialectal agent	Mean	20.13	.108	.089	.019	3.63	3.41	.38
	N	15	12	12	12	15	15	12
	Std. Deviation	6.82	.09	.08	.13	.53	.70	.21
Total	Mean	21.40	.103	.091	.012	3.74	3.34	.40
	N	30	27	27	27	30	30	27
	Std. Deviation	7.17	.10	.07	.13	.61	.75	.26

Figure 6.2: Student factors by condition

Independent samples t-tests revealed that there are no significant differences by condition in

terms of students' pre-test scores, dialect use, reading scores, or dialect attitudes.

From prior literature about students' dialect attitudes [Godley and Escher, 2012], dialect use [Charity et al., 2004b, Craig and Washington, 2004], and academic performance [Ogbu, 1999], I expected that the higher students' reading score, the more likely they would be to demonstrate either code-switching or reduced AAE use during the science pre-test, and the more positive attitudes they would have about SAE. Our correlation analyses of these features revealed that the only significant relationship was between students' attitudes about SAE and code-switching. The more students report positive attitudes about SAE, the more likely they are to demonstrate code-switching between the picture description task and the science pre-test. See figure x.x for the correlation matrix of each of these variables.

		Correlations					
		Developmental Reading Assessment score	Dialect (science pre-test)	Code-switching at pre-test	Ideologies about Standard English (pre-test)	Ideologies about Dialect (pre-test)	Reasoning with Science Discourse (pre-test)
Developmental Reading Assessment score	Pearson Correlation	1	-.063	.240	.101	-.030	.249
	Sig. (2-tailed)		.756	.229	.594	.875	.211
	N	30	27	27	30	30	27
Dialect (science pre-test)	Pearson Correlation	-.063	1	-.690**	-.189	.166	-.131
	Sig. (2-tailed)	.756		.000	.346	.407	.513
	N	27	27	27	27	27	27
Code-switching at pre-test	Pearson Correlation	.240	-.690**	1	.423*	-.080	.064
	Sig. (2-tailed)	.229	.000		.028	.692	.750
	N	27	27	27	27	27	27
Ideologies about Standard English (pre-test)	Pearson Correlation	.101	-.189	.423*	1	-.106	-.182
	Sig. (2-tailed)	.594	.346	.028		.577	.364
	N	30	27	27	30	30	27
Ideologies about Dialect (pre-test)	Pearson Correlation	-.030	.166	-.080	-.106	1	.043
	Sig. (2-tailed)	.875	.407	.692	.577		.833
	N	30	27	27	30	30	27
Reasoning with Science Discourse (pre-test)	Pearson Correlation	.249	-.131	.064	-.182	.043	1
	Sig. (2-tailed)	.211	.513	.750	.364	.833	
	N	27	27	27	27	27	27

Figure 6.3: Relationship between hypothesized student variables

There were no significant differences in the durations of students' interactions with the agent overall, nor during either the peer or teacher sub-tasks ( $p > .05$ ). The mean duration for students with the monodialectal agent was 20.03 minutes ( $SD = 2.6$ ) and the mean duration for students

with the bidialectal agent was 20.4 minutes (SD = 1.5). Though every interaction aimed to be 20 minutes long, differences in student behaviors or school logistics occasionally altered duration. One student with the SAE agent kept asking the agent if she could leave, for example, and she ultimately walked out of the room a few minutes prior to the experiment officially ending.

Due to extensive pilot testing and iterating of Alex's dialogue options, the agent was able to address many of the ideas students brought to the table that were more *off-script*. For example, one such unexpected student question was "*What is the creepy monster is just a big monster who can destroy all the people who come near it?*" In response, Alex countered "*Nuh-uh! He doesn't even have sharp teeth or claws or anything!*" The student responded "*Yeah, okay, that makes sense. Good point.*" Alex responded, "*So, then probably - I wonder if he can just eat the worms because he isn't very sharp?*" The student responded, "*Yeah, that makes sense, because he doesn't have sharp teeth.*" The interaction was then *back-on-track* without explicit redirection from the agent.

Occasionally, however, students were less easy to guide back into scripted dialogues. Here is one such exchange from a student in the monodialectal SAE condition.

*Tiera*: "Do you think the creature lives in the rainforest or the swamp?"

*Alex*: "Hmm, it looks like he's a really good climber because of his long legs, so... I think he lives in the trees. You?"

*Tiera*: "That's not what I said. I said rainforest or swamp?"

*Alex*: "I think the first one because there are so many trees there for him to climb on to help him protect himself."

*Tiera*: "Okay but Alex you're not listening. You have to say *the rainforest*."

At this point, Tiera was visibly frustrated. Looking at thin-slice rapport annotations, the slice prior to this exchange was labeled a 3 (slightly below neutral rapport) and this one was labeled a 1 (the lowest rapport label). Though our prior pilot testing had confirmed to us that we needed "filler answers" such as *the first one, that one, the second one, the last one* and so forth as a way

to address questions such as these, Alex was unable to provide the words Tiera wanted to feel satisfied with the interaction. After that last prompt from Tiera to *say the words* explicitly, Alex responded with one of the utterances designed to admit inability to comply. We worked with our voice actress to ensure that these utterances from Alex were able to seem mildly embarrassed, so as to promote forgiveness from the student. In this case, Alex responded *"I'm sorry... I don't really know how to talk about that stuff. What do you think about it?"* The student rolled her eyes and sighed audibly, and did not answer the prompt.

## 6.3 Results

### 6.3.1 Post-test science performance

Across all conditions, within-subjects t-tests demonstrated that there was a significant increase in students' use of science discourse from pre-test ( $M = 2.5$ ,  $SD = 1.6$ ) to post-test ( $M = 3.8$ ,  $SD = 2.3$ ), ( $t(25) = -2.92$ ,  $p = .007$ ). There was also a significant within-subjects increase across conditions of students' use of science argumentation from pre-test ( $M = .4$ ,  $SD = .26$ ) to post-test ( $M = .57$ ,  $SD = .29$ ), ( $t(25) = -2.1$ , ( $p = .048$ )).

There were significant differences in students' demonstration of science discourse over the course of the questioning,  $t(27) = -2.46$ ,  $p = .021$ ). Students' with the monodialectal agent demonstrated fewer instances of science discourse at post-test ( $M = 2.9$ ,  $SD = 2.2$ ) than those with the bidialectal agent ( $M = 4.7$ ,  $SD = 1.9$ ). Our analyses reveal differences by condition regarding the likelihood of students' use of science argumentation at post-test as well,  $t(27) = -3.08$ ,  $p = .005$ . Students' with the monodialectal agent demonstrated science discourse in approximately 45% of their reasoning responses ( $SD = .23$ ), and those with the bidialectal agent demonstrated science discourse during approximately 75% of their reasoning responses ( $SD = .3$ ).

### 6.3.2 Intervention science and rapport

In this section, we evaluate student utterances *during* the dialogue with our virtual agent, in sections of the task where Alex is guiding the discussion. These utterances occur as part of the intervention, prior to the administration of the post-test.

#### Intervention rapport

Across conditions, there was a positive correlation between students' average rapport with Alex throughout the session and the amount of science discourse they demonstrated at pre-test ( $r = .38, p = .048$ ).

Rapport was inversely correlated with social challenges ( $r = -.64, p = .002$ ). This was in line with work by [Creed and Beale, 2008, De Angeli and Carpenter, 2005], who argued that in their analyses of their own data, these social challenges appeared to be “psychologically aggressive” demonstrations of asserting their own power over the agent.

We began by assessing differences in students' average rapport by condition overall, as well as during the peer and teacher tasks independently. Across the full session, there was a significant difference in students' average rapport with Alex by condition over the course of the intervention. Students with the monodialectal agent demonstrated an average rapport of 3.61 ( $SD = .96$ ). This was significantly lower than students who worked with the bidialectal agent, who demonstrated an average rapport of 4.4 ( $SD = .86$ ),  $t(28) = -2.4, p < .05$ . We note that on average, students' rapport by condition generally centered around 4, which was the anchor of *neutral*. During the teacher formal presentation task, when Alex spoke SAE in both conditions, those with the monodialectal agent demonstrated an average rapport of 3.4 (slightly below neutral) and those with the bidialectal agent demonstrated an average rapport of 4.3 (slightly above neutral).

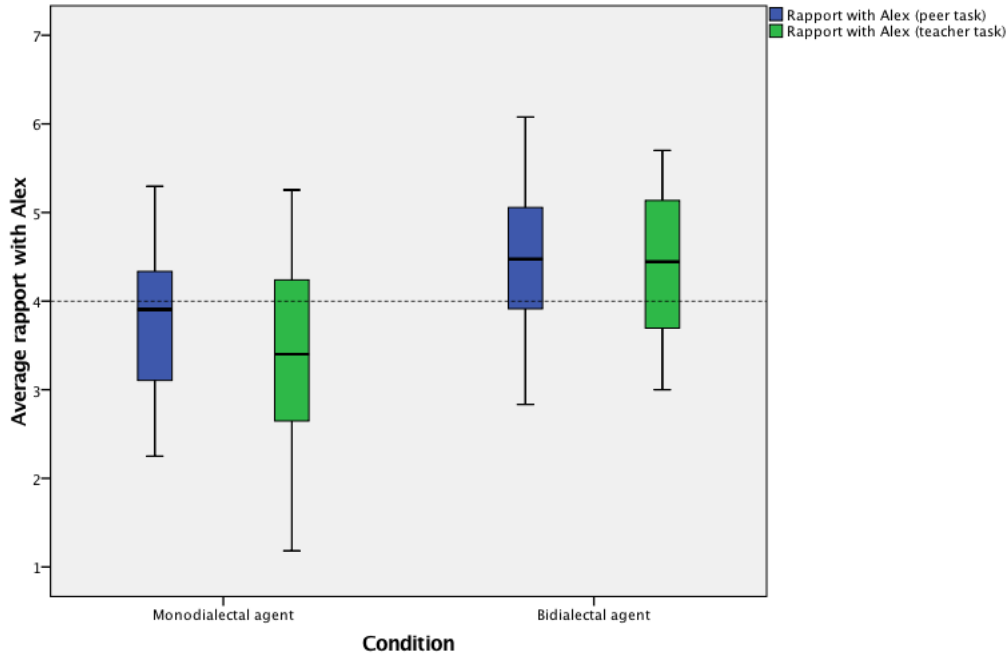


Figure 6.4: The impact of condition on intervention rapport

### Intervention science

Independent t-tests revealed that there were no significant differences by condition on students' demonstration of on-task talk, science discourse, or science argumentation.

### 6.3.3 Dialect density

As described in the beginning of this thesis, there are several studies that demonstrate integrating students' most comfortable language style into curricula can increase their demonstration of SAE during expected contexts [Wheeler and Swords, 2004, Sweetland, 2008, Fogel and Ehri, 2000]. In those interventions, AAE integration was primarily used in a technique called *contrastive analysis*, which explicitly compared AAE and SAE grammatical forms to highlight the differences. In the bidialectal condition of this intervention, we did not explicitly teach the distinction between AAE and SAE, but Alex modeled switching from AAE to SAE based on task context. As motivated in our first experiment using a distant peer “worked example,” we thus ex-

pected that Alex’s dialect pattern would likely not help students acquire *new* dialect patterns, but rather might be able to *elicit students’ code switching between dialects* if they already possessed the metalinguistic awareness necessary to follow such a model. In other words, we expected that Alex’s dialect pattern might *elicit* differences in students’ dialect use, either through a conscious decision to follow peer dialect norms [Eckert and Rickford, 2001, Alim et al., 2008] or perhaps an unconscious accommodation Alex’s speech [Giles and Marlow, 2011].

We calculated students’ percentage of AAE using a dialect density measure [Craig et al., 1998] as described in Chapter 3. For each student, we assessed students’ DDM throughout the whole session, during the peer task, and during the teacher task.

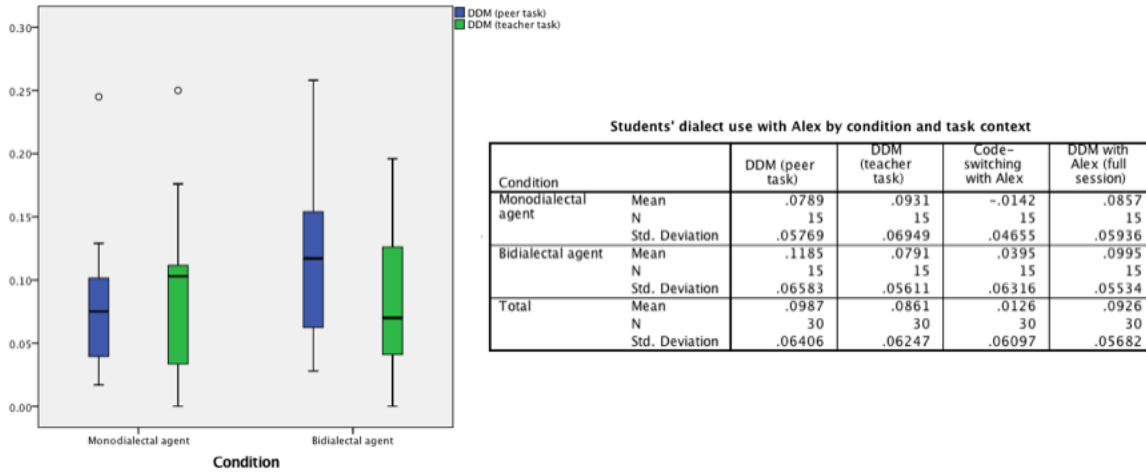


Figure 6.5: Dialect use with Alex by condition and task context

We performed independent samples t-tests to understand if there would be differences in students’ dialect use based on Alex’s dialect patterns. While students with the bidialectal agent trended towards using more AAE in the peer task and less AAE in the teacher task, these students did not demonstrate significantly different amounts of dialect overall from students with the monodialectal agent. However, there was a significant difference in students’ *code-switching between tasks*,  $t(28) = .61$ ,  $p = .013$ . In other words, students working with the monodialectal agent used more consistent dialect regardless of task context. Conversely, those with the bidialectal



tal agent demonstrated more context-specific dialect use, which is similar to the type of dialect shifting Alex demonstrated. We emphasize that students with the bidialectal agent didn't use less dialect during the teacher presentation task than those with the monodialectal agent, but that they demonstrated a greater *decrease* in the amount of dialect they used during the peer task.

Evidence also suggests that task context [Craig and Washington, 2004], idea complexity [Charity et al., 2004b], formality perception [Godley and Minnici, 2008], or interlocutor [Alim, 2006] might impact students' dialect use. To understand the relationship between Alex's dialect pattern and students' own contextual basis for dialect shifting, we examined the potential differences in code-switching between on-task and off-task talk within both task contexts by condition. In other words, we examined if students would demonstrate different dialect shifting behaviors between the peer and teacher tasks based on whether they were making an on-task contribution (e.g., "I think the creature eats worms") or an off-task comment (e.g., "Do you go to a school too?") We hypothesized that students with the bidialectal agent would demonstrate more code-switching of on-task comments, but that off-task contributions may not follow this same switching pattern. If this were the case, it might indicate that students weren't just mimicking Alex's dialect use directly, but rather were explicitly shifting their dialect for the on-task parts of the task.

We performed independent-samples t-tests on students' amount of code-switching between the peer and teacher tasks for their on-task speech and for their off-task speech. The data reveal that students with the bidialectal agent demonstrate significant within-subjects code-switching of their on-topic talk ( $t(14) = 3.36, p = .005$ ), but not their off-task talk ( $p > .05$ ). There is no significant code-switching within the monodialectal condition for either on or off-task talk. These dialect differences are shown in figure x.x below. However, the data reveal that students' DDM were generally relatively low for off-task talk across conditions. This may perhaps be attributed to the fact that these sentences were generally less syntactically complicated (e.g., "where's your teacher?" or "are we done yet?" than on-topic contributions. Even though the DDM measure we used accounts for SAE features by assessing the percentage of grammatical structures that

use AAE, this may not account for other context variables that impact dialect use such as speech genre [Craig and Washington, 2004] or idea complexity [Charity et al., 2004b].

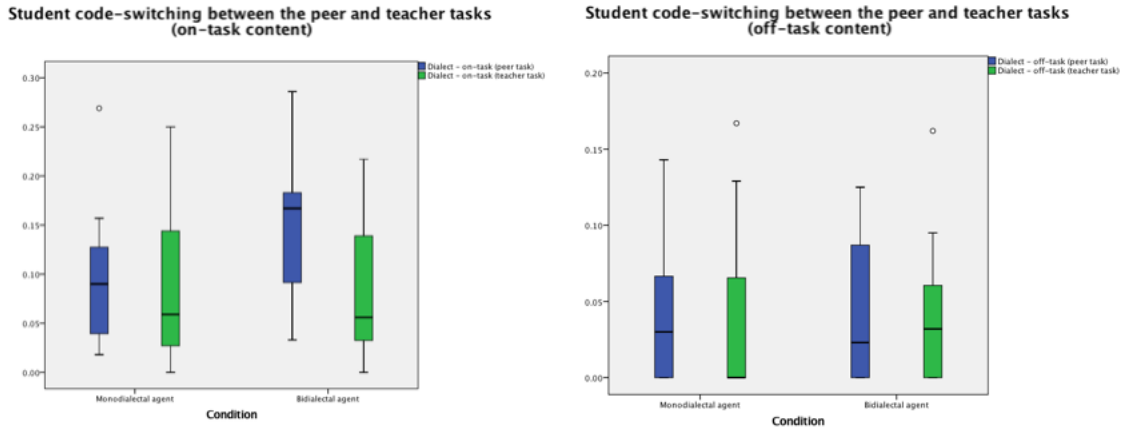


Figure 6.6: Code-switching with Alex by condition and task contexts

A common concern teachers' have against dialect-integrated pedagogy is the possibility that it will *increase* students use of AAE in other task contexts. As demonstrated in figure x.x, however, there was no difference in students' dialect use during the science pre- and post-tests by condition. We also performed paired-samples t-tests by condition to investigate if the intervention had an impact on students' within-subjects dialect use. There is not a significant increase in students' demonstration of AAE at post-test for students in the bidialectal condition. Similarly, there is also not a significant *decrease* in students' demonstration of AAE in the monodialectal condition, indicating that more exposure to SAE does not necessarily contribute to students' dialect use without additional instruction.

These analyses demonstrate that students seem to demonstrate more code switching between the peer and teacher tasks with Alex when they are paired with the bidialectal agent. Importantly however, they do not actually use significantly less AAE than those with the monodialectal agent. Looking at the data more closely, it appears that students with the bidialectal agent use slightly more AAE during the informal peer task and slightly less AAE during the formal teacher task, compared to those with the monodialectal agent whose dialect stays consistent across both tasks.

## 6.4 Secondary analyses to guide future investigations

I additionally present a section of secondary analyses where I aimed to investigate how three hypothesized variables might *moderate* the amount of impact Alex’s dialect had on student performance. I performed a backwards stepwise linear regression within each condition to identify how students’ dialect use, dialect attitudes, and standardized reading level impact their performance. However, due to our limited sample of students and the exploratory nature of that investigation, I aim for these results to serve primarily as recommendations for future work to investigate more directly.

### 6.4.1 Standardized reading level

Within the bidialectal condition, standardized reading level had no impact on students’ intervention rapport and science discourse. In both of these areas, students reading below grade level were indistinguishable from students at or above grade level.

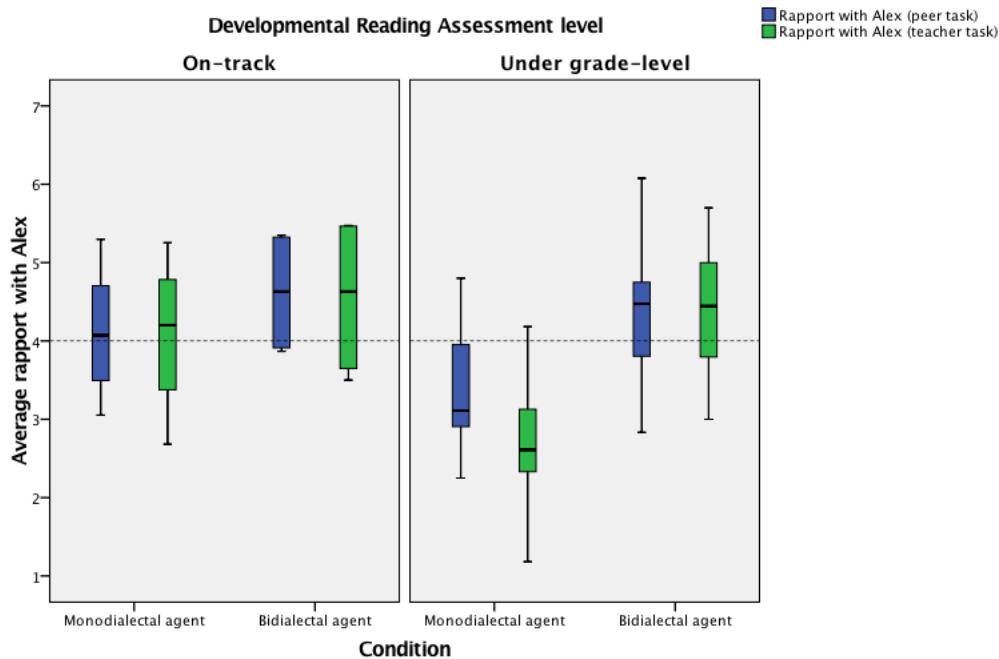


Figure 6.7: The impact of reading level and condition on intervention rapport

Conversely, with a monodialectal agent, there was a strong moderating impact of standardized reading level on students demonstrated rapport with the agent ( $F(1, 13) = 6.1$ ,  $p = .028$ ,  $\beta = -.565$ ). We confirmed using independent samples t-tests that there is no significant difference in rapport by condition among students who are reading at or above expected grade level. Conversely, within students who are reading under grade level, the impact of condition on rapport is considered a very large effect, Cohen's  $d = 1.8$ . This difference is visualized in figure x.x.

In other words, students who are considered on-track by their school's standardized metric of grade-level performance demonstrate equal levels of rapport with the agent regardless of Alex's dialect use. Conversely, students who are considered to be reading *below* grade level by standardized metrics within their school demonstrate a large difference in rapport based on the agent's dialect. When students reading below grade level worked with the bidialectal agent, they demonstrated rapport just as high as those students who are reading at grade level. Their counterparts who were randomly assigned to the monodialectal SAE agent, however, demonstrated substantially lower rapport than any other group of students.

This drop in rapport may explain why rapport during the teacher task was identified as a significant mediating variable driving the relationship between agent dialect and students' post-test science performance. It appears that students who demonstrated this rapport drop were also the students least likely to demonstrate learning gains at post-test.

In the prior section, we identified differences in students' post-test science argumentation by condition. Looking at students' post-test performance by reading level demonstrate an interesting distinction: it appears that students who report at or above grade level reading do in fact demonstrate improvements to their science performance at post-test. However, students reading below grade-level demonstrate no such gains. In the graph below I demonstrate these findings for both students' science discourse and science reasoning.

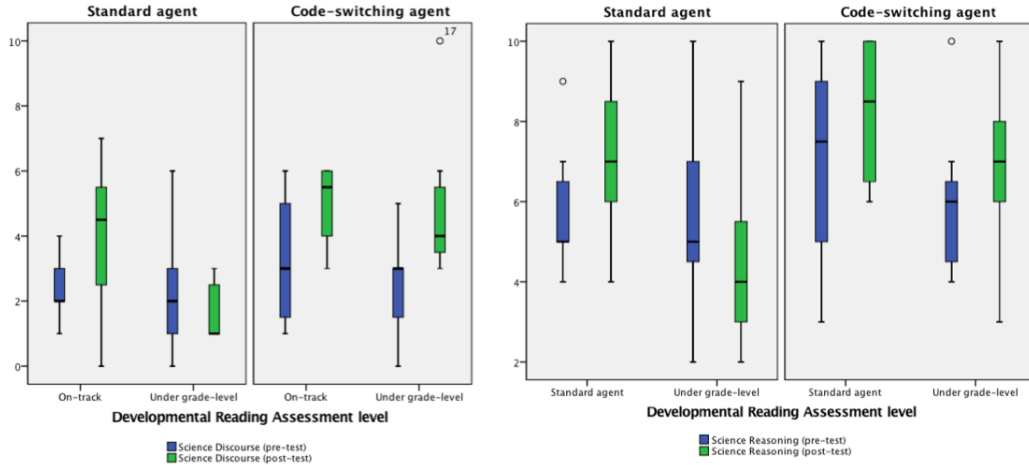


Figure 6.8: The impact of reading level and condition on intervention rapport

## 6.4.2 Dialect attitudes

Within the bidialectal condition, students' attitudes about Standard English emerged as a significant predictor of their use of science discourse with the agent ( $F(2,9) = 7.4, p = .01, \beta = -.63$ ). In other words, in the bidialectal condition, students demonstrated *more science discourse with Alex* when they reported *less positive attitudes about Standard English* at pre-test. This was in line with our hypotheses about the role dialect attitudes might play in students' science performance with the bidialectal agent.

To explore the potential impact of students' attitudes about SAE by condition, we used a linear model that took in condition, attitudes about SAE, and the interaction variable as predictors of students' use of science discourse with Alex. This model demonstrated a significant interaction between these variables. In other words, as figure x.x demonstrates, students' attitudes about SAE have different impacts on students' science discourse based on the dialect pattern of the agent.

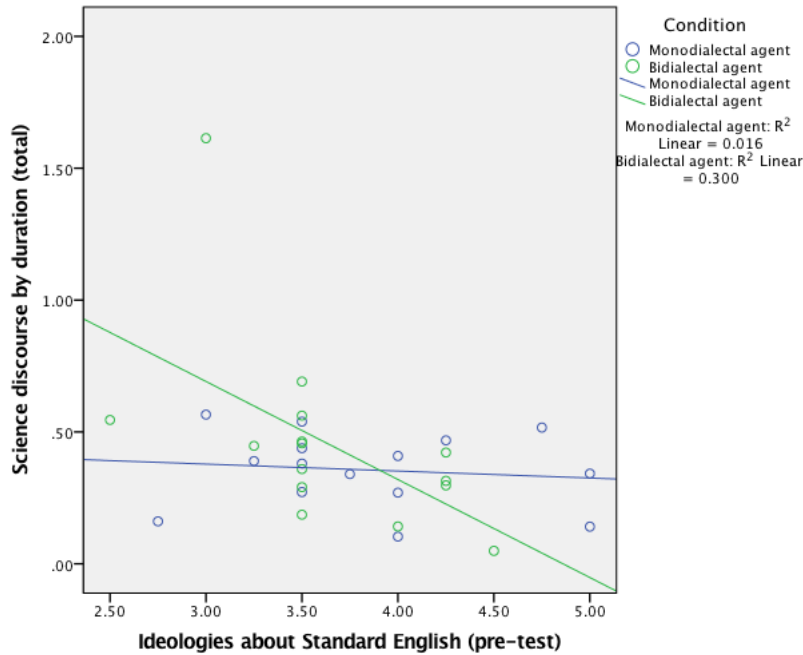


Figure 6.9: The relationship between attitudes about SAE and science discourse by condition

### 6.4.3 Dialect density

Students' dialect use during the science pre-test significantly predicted students' science discourse with the agent. This was a significantly predictive variable within the monodialectal condition ( $(F(1, 13) = 7.8)$ ,  $p = .02$ ,  $\beta = .6$ ), as well as across all students regardless of condition ( $f(17) = 6.69$ ,  $p = .008$ ). We performed a follow up analysis using a univariate linear model to investigate if there was an interaction between condition, dialect pre-test, and students' science discourse with the agent. This model demonstrated that there was no significant interaction between dialect use and condition. In other words, regardless of agent dialect, students who demonstrated *more AAE during the science pre-test* also demonstrated *more science discourse* with Alex.

## 6.5 Discussion

The results above adduce evidence for the argument that agent dialect may play a moderating role on students' experiences during an intervention in a way that impacts the level to which students demonstrate learning at post-test. Student science discourse and argumentation significantly increased in the bidialectal condition. This is not accompanied by a greater count of science discourse during the intervention itself; rather, student science discourse in the intervention is not significantly different by condition. Rapport, on the other hand, significantly differs by condition. This rapport increase is accompanied by significantly increased amounts of code-switching by students, while students dialect use *during pre- and post-tests* did not differ. This result is inconsistent with the claims that have historically been used as supports for exclusively monolingual classroom instruction.

### 6.5.1 Exploratory analysis discussion

Our exploratory analyses involving our hypothesized student factors revealed two potential relationships that could be explored more closely in subsequent work. Perhaps most importantly, it appears that agent dialect may not be a saliently impacting factor on students' performance when they are reading at or above grade level, one potential mark of on-track academic success. However, the majority of students in our sample were reading *under* grade level; for these students, agent dialect appears to have an impact on the intervention's efficacy.

In other words, these data suggest that for students who are already demonstrating success in at least one standardized area (reading), technologies that use AAE have no added benefit. On the other hand, for these students who are *already falling behind in school*, the difference is substantial. I also note again here that no part of this intervention (including the pre and post assessments) required reading. For this reason, it is likely reading level acted as a proxy for either a generalized measure of academic performance, or perhaps a representation of the complex social factors that may be associated with under-performance, such as an increased

disidentification from traditional academic experiences.

Students' *attitudes about SAE* seemed to play a different role in students' science discourse use based on Alex's dialect. In the monodialectal condition where Alex exclusively used Standard English, as would be the case in the majority of educational interventions, students' attitudes about SAE had no impact on their science performance with the agent. Conversely, for students in the bidialectal condition, there was a strong moderating impact of dialect attitudes. These analyses revealed an inverse relationship between students' dialect attitudes about SAE and their science discourse with Alex. In other words, the students who demonstrated *more negative attitudes about SAE at pre-test* were *especially likely* to demonstrate higher levels of science discourse with Alex. As was reported in a prior section, higher levels of science discourse with Alex was associated with higher levels of science discourse at post-test the next day. At the same time, students who felt more *positively* about Standard English at pre-test were *less* likely to demonstrate science discourse with bidialectal Alex.

These analyses also revealed that students' dialect use during the science pre-test may have also had a positive impact on students' science discourse with Alex, regardless of the agent's dialect. One explanation for this may be within the findings presented by several scholars described in Section 2.3.1, who find that students use of AAE may be *positively* predictive of their educational achievement as long as these students are also familiar with SAE. It is possible that these students *do* know SAE more fluently than they demonstrated here, and that they just did not view the pre-test as requiring of its use [Godley and Minnici, 2008]. Another potential explanation is that students who use more AAE during the science pre-test are students that are exhibiting less personal policing of their own dialect use. Some scholars posit that another way in which students may be negatively impacted by lack of SAE fluency is due to the cognitive work of trying to manage their own dialect production. If this interpretation were true, it might indicate that those students who demonstrate less AAE at pre-test are exhibiting fewer cognitive resources for dialect management, and in turn, demonstrate science discourse with the agent



more freely. It is also possible, of course, as with all of the results derived from these secondary analyses, that this is a false-positive result due to the running of multiple analyses.

### 6.5.2 Open questions

Though it is difficult to identify from these data why exactly some students may have demonstrated decreased rapport in the teacher task, we posit that there may also be several “*less exciting*” interpretations. I present these potential *alternative* explanations below.

The *innovation hypothesis* posits that the social reactions towards a computer are a temporary phenomenon due to the novelty of the situation. This novelty effect vanishes once the user becomes accustomed to the interaction with the technology [Kiesler and Sproull, 1997]. It is possible that the differences in social behavior observed in this study were simply due to the possibility that *science dialogues are just not fun*. It may be the case that the novelty of Alex’s dialect in the bidialectal condition mitigated the perhaps inherently *boring* nature of the task. This would also be one potential explanation for why students who were demonstrating at- or above grade-level performance demonstrated more rapport on average regardless of condition. It may be the case that the intervention was *fun enough* for students who were on-track in school. Conversely, for under-performing students, a bidialectal agent (whether due to its “taboo” nature, or simply its difference from the norm) might have just been better at *keeping students interested*. If this were the case, we would expect that over time, under-performing students would get *desensitized* to Alex’s bidialectalism, and demonstrate equally-low-rapport as those students with the monodialectal agent.

Another potential explanatory factor could be the *parasocial relationship* hypothesis [Calvert et al., 2014], which posits that students will be more likely to learn from virtual characters who feel familiar to them. It is possible that Alex generated a sense of *homophily* among those in the bidialectal condition due to the perceived similarity of dialect. Alim [Alim et al., 2008] posits that some students use AAE as a way of demonstrating a sense of community; such perceived *in-*

*group language* may have hastened the rapport building for students in the bidialectal condition, while those with the monodialectal agent would simply need more exposure in order to *catch up* with said rapport building. In other words, it's possible that Alex's bidialectalism generated a sense of familiarity, which promoted students' sense of sameness and, thus, rapport. If this were the case, we might expect that with increased exposure and thus familiarity, those with the monodialectal agent would *improve* their rapport with Alex over time.

Either of these two interpretations could be applied to both this study as well as the worked example study presented in Chapter 4. If these explanations accounted for differences in students' performance by condition, we would expect that over time, differences in performance by dialect would dissipate. This is an important factor to understand. If the findings represented by dialect integration dissipate with time, then these techniques might be useful for short-term implementation, but not as a broad factor to consider in a large-scale way. On the other hand, we posit that if these differences *do* maintain over time, dialect integration may be an important factor to consider moving forward in the design of educational technologies that promote equity. For this reason, we present next a longitudinal study that followed students' interactions with the agent over six weeks.

## Chapter 7

### Study 3: Longitudinal impacts

In this chapter, I present a longitudinal study we carried out to understand the potential impact of agent dialect over six weeks. This longitudinal investigation was carried out to provide more clarity on what impact agent dialect might have on student performance. By investigating the trend over time, we would be able to test for several alternative explanations as to what was driving the relationship between agent dialect and student performance. As described in the discussion of the above study, we posited that our results may have been attributed to either a *novelty effect* where those with the bidialectal agent would drop rapport over time, or a *familiarity effect* where increased exposure would improve the rapport among those in the monodialectal condition. In either case, we would then expect no differences in students' performances by condition after six weeks.

In this study, students interacted with Alex once a week for six weeks. Students worked with Alex individually on the three different science activities initially described in Chapter 5, and worked on each once over two sessions. Thus, weeks one and two were *the creature task* (the same activity described in the prior experiment), weeks three and four were *the bridge task*, and weeks five and six were *the ramp task*. The sessions each week largely took place identically to the paradigm described in the prior experiment.

Because of the limitations of the school's availability to host us in an isolated room, required

class activities students had to attend during the week, and our desire to work with each student once per week, we decided on a case study design where we worked with twelve students. Six students worked with Alex in each of the two dialect conditions. The twelve students in our study were given pre-tests, worked with Alex for six weeks, and then post-tests the following week. We additionally gave pre- and post-tests to eight other students who did not receive any intervention so that we could compare students' post-test performances to a control condition to account for time.

## 7.1 Research questions

As in our prior study, we investigate students' science discourse, dialect use, and social behaviors with Alex during each session. We investigate if these behaviors change over time, with a particular focus on understanding if any potential differences disappear after six weeks. We hypothesized that any convergence between conditions could be due to two explanations. One potential explanation for convergence could be the novelty effect wearing off and students' participation with the bidialectal agent decreasing over time. Another potential explanation for any convergence could be due to students with the monodialectal agent building up their rapport and participation with the agent over time. Such an explanation would be in line with explanations of rapport that argue that those with more initial coordination (perhaps due to similar dialects) might demonstrate a stronger *initial* rapport, but that coordination improves with repeat exposure [Tickle-Degnen and Rosenthal, 1990, Calvert et al., 2014].

## 7.2 Methods

This study was carried out at the same school described in the prior study. Though the prior study was done with second grade students, we worked with third grade students in this experiment due to limitations regarding participant access. Work from prior scholars that have studied grade

related changes in students' AAE use found a slight decrease in students' AAE use on average between 2nd grade (4% morphosyntactic AAE features) and 3rd grade (3.5% morphosyntactic AAE features). I note here that these authors normalized students' counts of AAE features over *number of student words*, whereas in our work, we normalized students' counts of AAE features over *counts of SAE features*. Regardless, we expected the 3rd grade students' own AAE use might be reduced in this study compared to the density of production from 2nd grade students in the prior study.

In this study, we also did not obtain parental consent to collect students' reading level scores, and thus this data is left out of our analysis. However, from our qualitative work with the students (e.g., certain students being called out of the classroom for particular events), we have some insights into students' academic standings in various ways. We include some of this information in our case studies of the results in the results section below.

There were two procedural differences between this study and the one described above. The first is Alex's priming of dialect use. In the last study, Alex told students in both conditions that "*My teacher likes it when I use my school English whenever I do presentation stuff like this.*" One hypothesis might have been that students in the monodialectal condition experienced a drop in rapport because Alex prompted them to use "*school English*," even though Alex was speaking SAE the whole time. In doing so, it may have been seen as *face-threatening*, or perhaps patronizing, to receive that type of linguistic prompt from someone who was not speaking within their same discourse community. In this study, Alex did not reference dialect in any capacity within the monodialectal SAE condition, and prompted the switch to SAE the same way as described in the prior study. Thus, in this study, the monodialectal condition involved Alex only speaking SAE and not referencing dialect at all, and in the bidialectal condition, Alex spoke AAE in the peer task, prompted a switch to SAE for the teacher task, and then only spoke SAE for the remainder of the session.

The second difference between the prior study and this one is that Alex prompted the students

to give a science presentation in their own dialogue at the conclusion of each of the three science activities (so at the end of the second, fourth, and sixth sessions). For every student, Alex first presented a model presentation that covered the main science arguments that Alex aimed to guide throughout the session. This science presentation was pre-recorded, and thus identical for every student. After Alex concluded, Alex invited the student to give a presentation of their own. Unlike the prior study, there were not specific question prompts given for the student to answer. Alex’s presentations lasted approximately 40 seconds, and students were able to speak for as long as they chose to on their own presentations.

## 7.3 Student case studies

Though we had a small number of participants, we were able to work with each student six times through their interactions with Alex. To better explain our results below and offer our interpretations, I begin this section by providing a brief qualitative overview of two students, one per condition, that were *extreme* examples of the representative differences we found in across this experiment. These two students each demonstrated salient behaviors with the agent that may help provide qualitative interpretations of our results. All names have been anonymized. I here present a short description of what we learned about this student through their dialogues with Alex.

### 7.3.1 Jurnei (bidialectal agent)

Jurnei demonstrated consistently high rapport with Alex throughout the six week study. Jurnei made several admissions over the course of her interactions with Alex that she doesn’t like paying attention in school, that she’s “*always in trouble*”, and that her teacher doesn’t like her. Her exchanges with Alex were also largely marked by their very social and communicative nature. Alex began each session with every student with a small amount of introductory talk (e.g., “how’s

it going?”). While most students responded with a variation of “*pretty good, you?*”, Jurnei had new insights about herself to share with Alex each week. During the morning of her 6th session with Alex, at the beginning of reviewing the ramp task, they had this exchange:

**Alex:** How’s it going?

**Jurnei:** Well I was upstairs at school bef– upstairs before I got down here, and we was setting goals for what we wanna be whenever we grow up

**Alex:** Oh! Cool.

**Jurnei:** I wanna be a basketball player.

**Alex:** I don’t play no sports this year, but last year I play basketball.

**Jurnei:** That’s what I wanna be when I grow up! A basketball player!

**Alex:** Cool! I like sports. I wish we had more time to talk about fun stuff.

**Jurnei:** I know, me too. So now we have to make all these ideas into our presentation now huh?

At this point in the dialogue, students outside the door began talking and their exchange was audible from within the enclosed room where the intervention was taking place. At this point, Jurnei stopped and addressed the students.

**Jurnei** (*looking towards the door*): Hey! Keep it down back there!

**Alex** (*ignoring the students outside the room*): So... you think we should start with the high one or the low one?

**Jurnei:** Wait, I’m sorry Alex, hold on a sec

**Jurnei** (*again addressing students outside the room*): I said keep it down out there! I’m tryin’ a do something with Alex!

**Alex:** Thanks!

**Jurnei:** It’s okay. Let’s do the high one first.

Jurnei also made several references about enjoying friendship with Alex over the course of the session. The following example comes from the beginning of her fourth session with Alex.

**Jurnei:** We was doin' the daily five right now, I just got off the computer.

**Alex:** What's that?

**Jurnei** (*smiling*): A computer? Alex! It's a computer!

**Alex:** It work different when you live in a computer world like me.

**Jurnei:** Yeah. I wish you could sit right here next to me right now so it would be easier.

**Alex:** Me too. That sounds fun.

**Jurnei:** We'da been best friends, like we are right now.

### 7.3.2 Terran (monodialectal agent)

Like Jurnei, Terran also made several references to not liking school. Conversely, Terran's references to not liking school often seemed to be situated in exchanges where Terran was also communicating an outward dislike of Alex. While Terran only demonstrated a few instances of aggression towards Alex on the first session, these behaviors increased over the course of the session. Terran demonstrated frequent displays of what we called *testing behaviors*, a type of social challenge that seemed to involve reminding Alex of his machine-nature or enforcing Alex into his "role" as a computer.

In prior work about agent abuse, De Angeli and Carpenter [De Angeli and Carpenter, 2005] explain that these types of behaviors demonstrated to computers often do not seem to show frustration the way one might show frustration to a stalled car; rather, he argues, these behaviors seem to demonstrate "psychologically sophisticated" aggressions that seem intended to cause harm. While I state again that it is not possible to understand what a student is feeling but simply their outwardly demonstrated behaviors, we are not able to fully interpret why students demonstrate the behaviors that they do. However, I report here salient behaviors demonstrated by Terran throughout his experiences with Alex. As I will describe further in the quantitative results below, these behaviors were often labeled as *social challenges*, and were negatively associated



with students' science performance.

Though Terran demonstrated several social challenges towards Alex during the first two sessions, these behaviors became particularly salient from the very beginning of the third session. The following is the very beginning of Alex and Terran's interaction during the third week.

**Alex:** Hey! How's it going?

**Terran:** Blaaaaah

**Alex:** What were you doing in class before you came here?

**Terran:** None of your beeswax.

**Alex:** Um, okay.

**Terran:** What were *you* doing in class, before you came *here*?

**Alex:** The rest of my class is doing reading now.

**Terran:** I don't see them.

**Alex:** My teacher pulled me out to do this.

**Terran:** I don't see them. I don't see them. Where are they?

**Alex:** My teacher pulled me out to do this.

**Terran:** Where are the students?

**Alex:** I think everyone else is doing math groups now.

**Terran:** Where are the students?

The above exchange demonstrates the occasional limits of Alex's dialogue options. The majority of the team, the initial responses Alex gave were enough to quell students' curiosities or concerns. However, in the instances where students began to continue asking as Terran demonstrates here, Alex failed to be able to successfully continue on the dialogue. In doing so, Alex's "machine state" becomes more immediately apparent. Over the course of Jurnei's six week session, Alex rarely or never was *unable to respond* to a question Jurnei asked. At one point, when she *did* ask a follow up question that Alex couldn't answer, Alex admitted "*I don't really know how to talk about that stuff.*" In return, Jurnei said "*That's okay Alex, let's just get back to the*

*science.*”

Terran however was often not as *lenient* as Jurnei. Resultingly, Terran managed to often bring Alex to dialogic states outside the bounds of which he could reasonably answer. When this happened, Terran effectively managed to *break* Alex’s ability to demonstrate fluency. As I described during the discussion of the prior study, some scholars argued that academic tasks have inherent power embedded within them [Cummins, 2000]. In a similar token, Fordham [Fordham, 1999] reports qualitative findings that African American youth may demonstrate *resistance behaviors* as a way of maintaining dignity when they are being asked to participate in an academic task that they do not value.

Terran’s interactions with Alex were marked by rejections of Alex’s ideas, refusal to do the task, and profane demands that inappropriate to repeat in this document. However, these rejections were often not as simply as simply *ignoring* the agent, but often involved complex demonstrations of tricks or traps that could be seen as efforts to catch Alex in a lie, or embarrass Alex. Terran also demonstrated several instances of linguistically sophisticated move called *format typing*.

Format typing is an intentional re-use of what has just been said in a dialogue, often for narrative or emphatic effect. This language behavior in particular has been identified as a linguistic strength of African American youth that has been observed as early as pre-school. For example, Corsaro [Corsaro and Eder, 1990] identifies that African American pre-school children in a head-start program would use format typing in their peer dialogues to a far greater extent than White American children from a different pre-school program. One such exchange Corsaro reports is a dialogue between two four year old children. The first says, “I don’t know what you’re laughin’ at.” The other responds, “I know what I’m laughin’ at, I’m laughin’ at your head.” I report below two more brief exchanges between Alex and Terran, both of which involve Terran’s use of format typing.

This exchange took place at the beginning of the fourth session with Alex during the peer

brainstorming task.

**Terran:** Do you know my name?

**Alex:** Yeah, I remember it from last time.

**Terran:** What's my name? What is it?

**Alex:** Yeah.

**Terran:** What's my name? Say it.

**Alex (*smiling and looking down*):** I'm not sayin'

**Terran:** Please say my name 'cause I don't think you know my name.

**Alex:** I wish we had more time to talk about fun stuff.

**Terran:** Say my name because you don't know my name if you don't say it.

**Alex:** I think we have to talk about the science stuff now.

**Terran:** I think we have to talk about *my name*.

About ten minute later, after Alex and Terran had switched to the teacher presentation task, this exchange took place.

**Terran:** Now I'm the teacher. How many bags fit on the bridge?

**Alex:** Hmmm... four.

**Terran:** Actually six.

**Alex:** Why?

**Terran:** I'm the teacher, you can't ask me why.

**Alex:** Okay. Can you teach me so I know it better for next time?

**Terran:** Uh like, you can fill up the bags for two more, like the top row and then the middle row.

**Alex (*elongated*):** Ooooh.

**Terran (*also elongated, matching Alex's prosody*):** Yeeaahh.

I present below one more set of excerpts from our data, again contrasting Jurnei and Terran. In doing so, I aim for these examples to help provide additional context for the quantitative results

I report in the remainder of this chapter.

### **7.3.3 “Alex, do you speak Spanish?”**

I provide one additional set of examples here to provide an understanding of the ways in which these sessions with Alex did not always go according to a pre-meditated plan. Students were taking a Spanish class over the duration of their participation in this intervention, and several students mentioned Spanish to Alex in various ways. I contrast here the ways that Jurnei, Terran, and K’Mya asked Alex about Spanish as a representation of these differences in their behaviors.

This interaction took place during the sixth and final session with Alex, as they were transitioning to the teacher task.

**Alex:** My teacher likes it whenever I use my school grammar whenever I have to do presentation stuff like this.

**Jurnei:** Yeah, my teacher does too. It makes it like – it makes you make more sense, like so the person can understand what you’re saying.

**Alex:** I can talk in different ways in different situations, like school English and dialect.

**Jurnei:** Oh – can you speak Spanish and French?

**Alex:** No, I didn’t learn that in school. What about you?

**Jurnei:** I used to, when I was in, um, first grade and kindergarten. The only Spanish I know is my numbers, for real- but I really –

**Alex:** That sounds fun.

**Jurnei:** You want to say ’em? Spanish numbers?

**Alex:** Yeah!

**Jurnei:** Uno, dos, \*cues, cuatro, cinco, seis, \*siente, ocho, nueve, diez

**Alex:** Cool - I like it!

**Jurnei:** Maybe you should learn ’em someday ’cause I know a lot of schools take

Spanish classes I've been hearing

**Alex:** Yeah, that makes sense.

This exchange, from Terran, also took place during the fourth session. It occurred within the peer brainstorming task. Alex and Terran were talking about how to improve the bridge's structural design so it would be stronger.

**Terran:** Fill it up for the middle.

**Alex:** Yeah, so then the beams can be really strong.

**Terran:** Do you know what this is? *Disculpe*.

**Alex:** Can you um teach it to me?

**Terran:** Si. You wanna learn Spanish? That's what I was talkin'

**Alex:** Oh! Cool.

**Terran:** Si. *Disculpe. Hombre. Mujer. Rojo.*

**Alex:** That sounds fun. I didn't learn about that in school.

**Terran:** Blah blah blah blah blah blah. Take some off. The blocks. Some blocks so it's even for the sides. Or add blocks.

**Alex:** Uh huh?

**Terran:** See the difference? Look at them, you're not even looking at them.

**Alex:** Do you think the red and green side beams should be um the same size as the rest of them, or-

**Terran:** You interrupt me. You interrupt me. You interrupt me. You interrupt me.

**Alex:** Sorry for interrupting!

**Terran:** You interrupt me. You didn't look at the green or red side.

**Alex:** It looks like the green side beam only has one row, and the red beam has two rows.

**Terran:** Ya think? Boy you're stupid, so stupid. So stupid.

**Alex (groaning):** Mmmmm...

**Terran:** Disculpe. Disculpe. Sí. Stupid. You're stupid. You're stupid. Stupid.

## 7.4 Results

Throughout the remainder of this section, I report our quantitative results of the variables we have addressed in our prior studies: students' science discourse, science argumentation, and dialect use. I present our broader interpretations of these results in our following discussion section.

Unlike in the prior study where students' science performance was assessed the next day, in this study students' gave their final presentations *to Alex* at the end of the second, fourth, and sixth session. In each of these presentation opportunities, Alex first modeled her presentation and then students were asked to present their own. This was still done under the guise of *practicing* a presentation they would later be able to give to a teacher. During the second, fourth, and sixth sessions, Alex introduced the final presentation task approximately 15 minutes into the session (five minutes into the teacher task).

I note here that one student who had been assigned to work with the monodialectal agent only participated in three of the six sessions (the first, fourth, and fifth) due to frequent long absences, at least one of which was due to suspension. We did not immediately realize this, as the classroom teacher we were working with had (likely accidentally) sent a different student in her place twice. This student had the same first name as the initial student, and had also consented to participate in the study. However, because this student had been assigned to the control condition rather than one of the agent conditions, she was not intended to work with the agent. We did not realize our mistake until the fifth session when our initial student returned back to class. Because of the nature of the analyses we did here aimed to understand change over time with increased exposure however, we ultimately do not include the results from either of these two students in our statistical analysis. Thus, the statistics presented below are from six students who worked with the bidialectal agent and five students who worked with the monodialectal agent.

### 7.4.1 Science discourse

The effect of condition from our first study is replicated in the longitudinal study. In a model including both session and condition, students in the bidialectal condition produce significantly more science discourse ( $f(20.4) = 7.1$ ,  $p = .015$ ) and more science reasoning ( $f(29.1) = 13.2$ ,  $p = .001$ ) than students in the monodialectal condition. There was no significant difference in the number of students' science discourse or reasoning across sessions or by condition when analyzed over the full 20 minute interaction.

We then investigated potential differences in students' science argumentation in both the peer tasks and teacher tasks over time separately. Within the peer task, there was no significant effect of either condition or session on students' science argument production. Within the teacher task, the linear model revealed a significant effect of session ( $p = .004$ ), though no significant effect of condition. We evaluated if this result would remain by excluding sessions two, four, and six, such that we were then only comparing students' use of argumentation during their first exposures to each new task.

I then looked at students' use of science argumentation during their *final presentations* to Alex which were presented within a five minute window at the end of the second, fourth, and sixth sessions. Repeated measures ANOVAs revealed no significant increases in students' uses of science reasoning, science discourse, or science argumentation over time over all, with no significant interaction of condition.

### 7.4.2 Rapport

I then aimed to investigate whether rapport played the same predictive role on students' science presentation discourse as we found in our previous study. I ran a linear mixed model that evaluated the results of condition, rapport, and an interaction variable for condition and rapport. These results revealed a significant interaction between condition and rapport on students' final presentation science discourse. The data reveal that rapport did not seem to play a predictive role

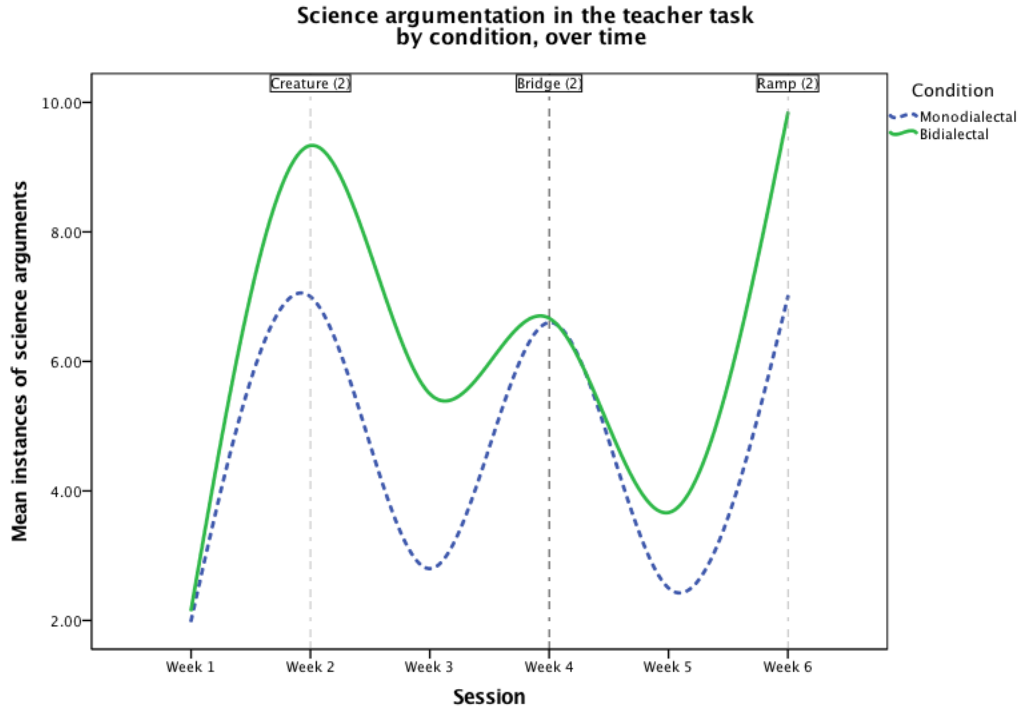


Figure 7.1: There was a significant impact of session on students’ science argument production which seems to be lead by students’ final presentations the 2nd, 4th, and 6th sessions. There is no significant impact of session after controlling for final presentations.

on students’ final presentation discourse, perhaps because the rapport within this condition was fairly consistent. However, like in the study presented in Chapter 6, it appears that rapport was far more variable among the five students with the monodialectal agent. For these students, rapport was strongly predictive of their final presentation discourse. I describe these results further in the following section.

### 7.4.3 Social behavior

We ran linear models that investigated session as a repeated measure, and investigated the impact of condition and session on students’ rapport with the agent in both the peer and teacher tasks. These models revealed that there was a significant difference in rapport by condition in the peer



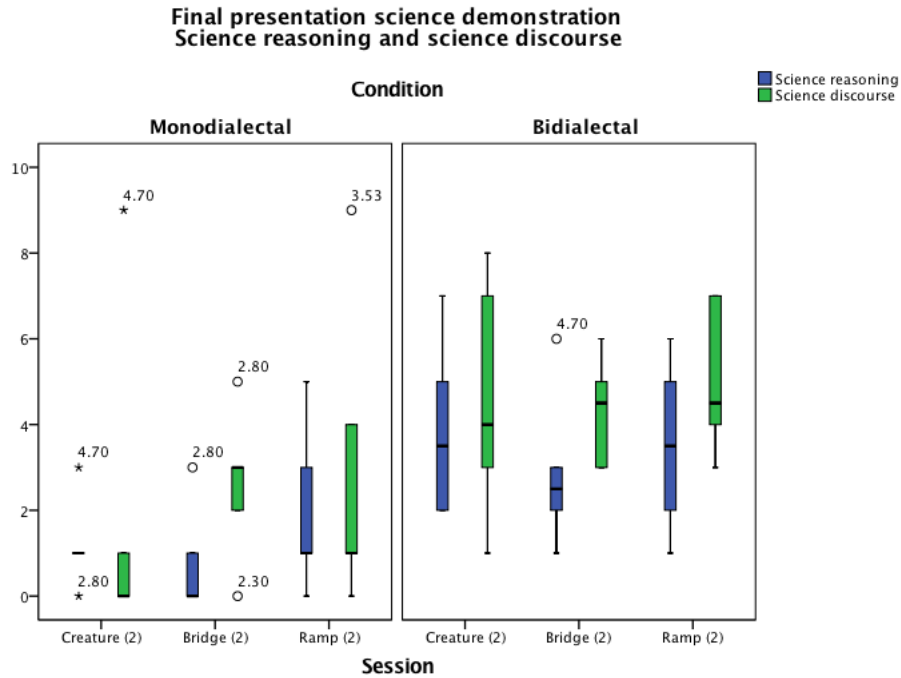


Figure 7.2: Science reasoning and science discourse during final presentations

task ( $f(49.3) = 4.12$ ,  $p = .047$ ) and in the teacher task ( $f(49.2) = 12.12$ ,  $p = .001$ ). These models revealed that there was no significant effect of session, where students generally demonstrated less rapport over time, regardless of condition.

We also assessed the amount of *positivity* and *negativity* in students' speech over time and by condition. As described in Chapter 5, positivity consisted of praise ("*I like your hair!*", sameness of character ("*I like basketball too!*"), and sameness of cognition ("*I was thinking that same thing!*"). Negativity consisted of social challenges ("*Wrong answer, stop talking*") and insults ("*You are so stupid.*") Linear models revealed that there was no significant difference in students' demonstrations of positivity by condition or over time. There *was* however a significant difference in students' demonstrations of negativity by condition (stats).

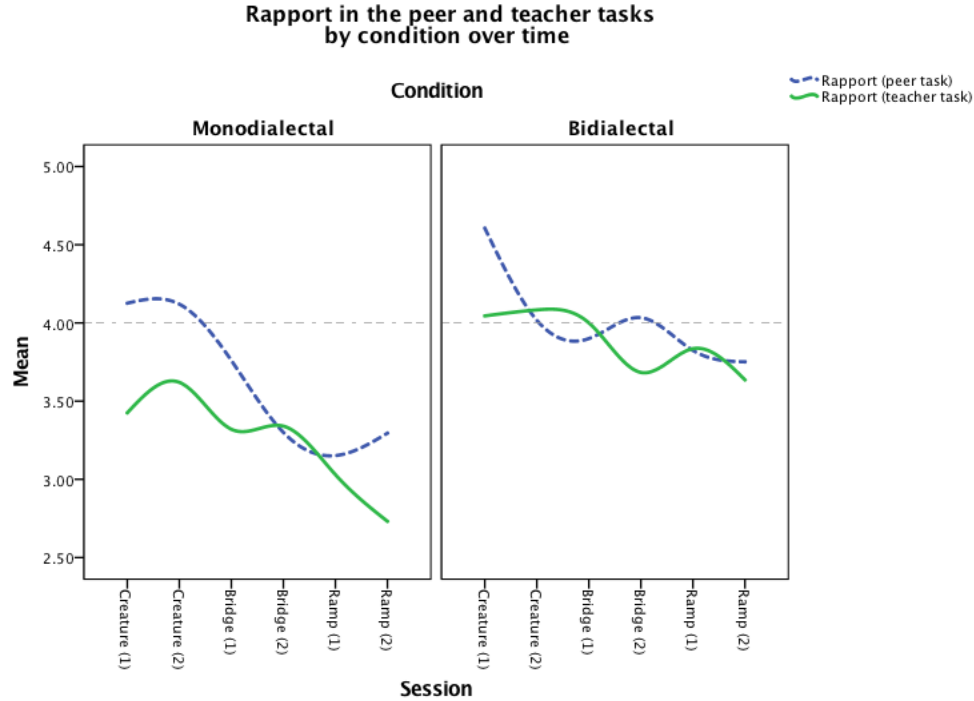


Figure 7.3: Rapport over time by condition

#### 7.4.4 Dialect density

We performed linear models to investigate the difference in students' AAE dialect density (DDM)s in the teacher task and the amount of code-switching demonstrated by students through the course of the intervention. These analyses revealed that students with the bidialectal agent used significantly more dialect during the teacher task than those with the monodialectal agent ( $f(48.6) = 4.9$ ,  $p = .032$ ). There were not significant differences by session.

Because there was no difference by session, we report here students' DDMs by task averaged over the six sessions by condition. Across all six sessions, students with the monodialectal agent demonstrated an average DDM of 8.3% in the peer task ( $SD = .08$ ) and 5.5% ( $SD = .04$ ) in the teacher task. Students with the bidialectal agent demonstrated an average DDM of 8.3% in the peer task ( $SD = .09$ ) and 9.3% in the teacher task ( $SD = .08$ ).

Across all six sessions, students with the monodialectal agent generally demonstrated a re-

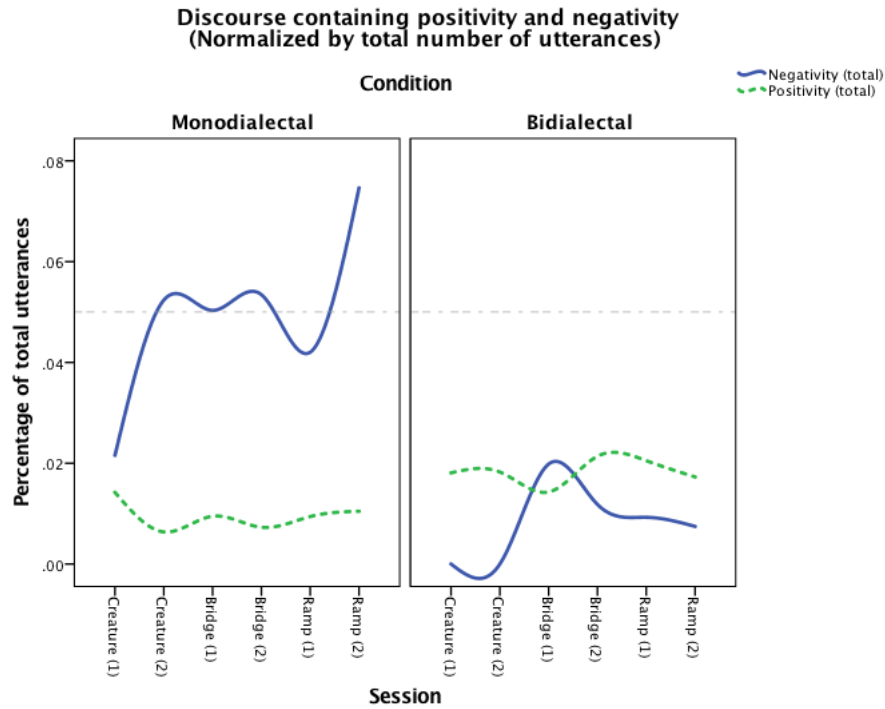


Figure 7.4: Percentage of student utterances containing positivity and negativity over time by condition

duction in their dialect use by 2.3% ( $SD = .08$ ). Those with the bidialectal agent demonstrated an average increase in their dialect use by 1.1% ( $SD = .13$ ) between the peer and teacher task. We performed a linear model to assess if there were significant differences by condition in the amount of code-switching demonstrate, and found that neither condition nor session were significant ( $p > .05$ ). In other words, though students with the bidialectal agent demonstrated significantly more AAE during the teacher task, there were not differences in the amount of code-switching between tasks completed by either group.

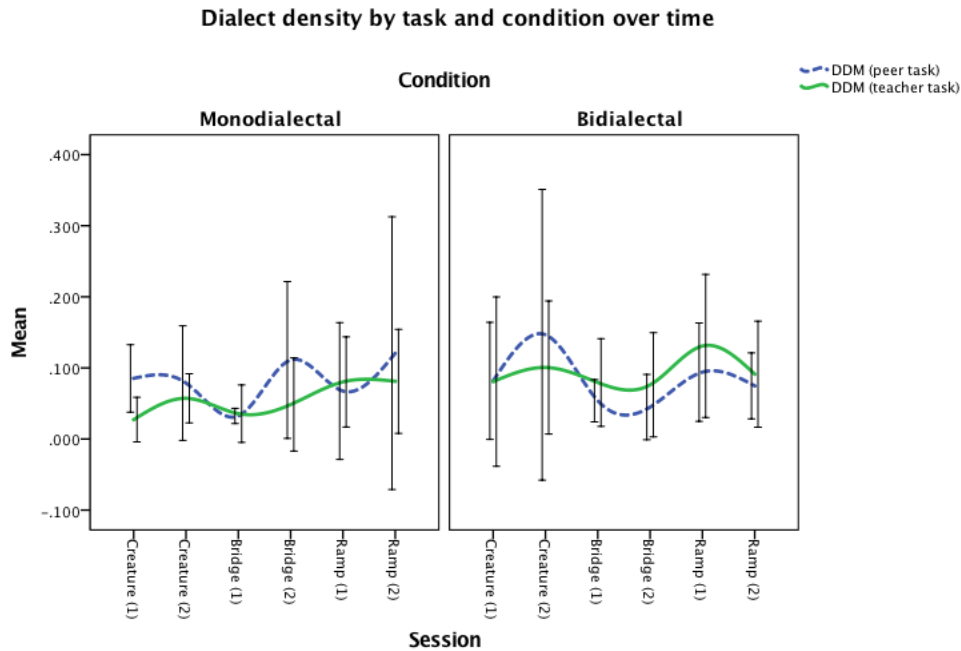


Figure 7.5: Students' dialect use between tasks and across sessions by condition

## 7.4.5 Language awareness and attitudes

### Language attitudes

We measured students' explicit and implicit attitudes about both SAE and AAE at pre-test and post-test.

First we report on students' attitudes about AAE by condition. There was no difference in students' explicit language ideologies about AAE at pre-test ( $f(2) = 2.116, p > .05$ ). At post-test, however, there was a significant difference between the three groups ( $f(2) = 7.208, p = .008$ ). Students in the bidialectal condition ( $M = 4.04, SD = .86$ ) reported significantly more positive attitudes about AAE than those in the monodialectal condition ( $M = 2.55, SD = .32$ ) and the baseline condition ( $M = 3.05, SD = .64$ ). There was no difference between the post-test results of the monodialectal and baseline conditions.

Additionally, a paired-samples  $t$ -test within the bidialectal condition demonstrated that students significantly *increased* their positive attitudes about AAE from pre- to post-test ( $t(5) =$

-3.588,  $p = .016$ . Conversely, those in the monodialectal condition demonstrated significantly *less positive* attitudes about AAE at post-test than they did at pre-test ( $t(4) = 3.413$ ,  $p = .027$ ). There was no change from pre- to post for the students in the baseline condition ( $p > .05$ ).

We here present students' *implicit* dialect attitudes. There was no significant difference at either pre- or post-test in students' assessment of speaker intelligence based on dialect ( $p .05$ ). However, across both conditions, there was a significant change in students' perceptions of the intelligence of SAE speakers compared to AAE speakers at both pre-test ( $t(14) = -4.62$ ,  $p < .0001$ ) and post-test ( $t(18) = -5.2$ ,  $p < .0001$ ). At both pre and post-test, students in all conditions rated the intelligence of SAE speakers substantially higher than AAE speakers ( $M = 4.2$ ,  $SD = .80$ ) than AAE speakers ( $M = 2.6$ ,  $SD = .91$ ).

## 7.5 Discussion

This study successfully replicated the main effect of condition from the first study of Alex. Students in a bidialectal condition produce more science discourse and reasoning than students in a monodialectal condition. Based on the repeated interactions in this new longitudinal setup, we then effectively eliminate null hypotheses based on novelty or familiarity that could explain this effect. Such hypotheses predict that the difference in student science discourse across conditions would change over time. This does not occur; science discourse stays consistent across conditions, with no effect of session.

By contrast, the amount of science discourse students demonstrate is more closely predicted by the rapport they demonstrated in that session. As students gained familiarity with Alex over time, they do not grow in rapport; in fact, rapport declines. The impact of rapport by condition appears to be driven by social challenges, a behavior which appears disproportionately in the monodialectal condition. In the context of virtual agents, this abusive behavior has a negative effect on learning, predicting reduced science performance.

There is likely an introductory novelty or familiarity effect on rapport between students and

virtual agents. However, this is insufficient to explain rapport differences by condition; after six weeks, the difference remains significant. Given the data that is available from this study, it is not possible to estimate whether conditions would become equal after more than six sessions.

Student attitudes about, and production of, AAE also differed significantly across condition and were not subject to novelty or familiarity effects. Students in the bidialectal condition produced more AAE language while maintaining similar levels of science discourse and reasoning compared to the baseline, and their attitudes on AAE became significantly more positive than the attitudes observed in the monodialectal condition. Despite this, students in all conditions continue to rate SAE as a substantially higher-intelligence dialect.

These results confirm that some effects of bidialectal agents are consistent across sessions and over time, and that novelty or familiarity effects are insufficient hypotheses. The impact of condition appears to be mediated by rapport, and in particular, by abusive social challenges in the monodialectal condition. In the next chapter, we discuss the potential ramifications of these results.

## Chapter 8

### Research discussion

In each of the three studies presented in this thesis, we found an impact of our experimental manipulation on students' production of science discourse. First, in each of the three studies, there were *within-subjects differences* between students' use of science discourse from pre-test to post-test regardless of condition. Next, students benefited from a bidialectal pedagogy. Students in both conditions demonstrated more science discourse; students who worked with a bidialectal agent demonstrated *more* science discourse. This confirms the goals of both of our primary questions for this thesis.

An improvement in rapport with Alex, varying by condition, is an explanation of these results supported by the data. Students built more rapport with Alex in the bidialectal condition; then, rapport was positively correlated with the amount of science discourse they demonstrated at post-test. This is best understood through the lens of social challenges, which are more common in the monodialectal condition. The bidialectal condition also has an impact on student attitudes about AAE; students in this condition used more AAE grammatical features and held more positive beliefs about the dialect.

This result persists across time. In the final, longitudinal study that we present, the impact of the bidialectal condition on science discourse was replicated and shown to persist across six sessions. The downward trend of rapport over time was not affected by condition. This rules

out novelty or familiarity effects as a simpler explanation for the results observed in the transfer study. These results, both collectively and individually, are consistent with our research hypotheses on the impact of bidialectal pedagogy in the context of a pedagogical agent.

## 8.1 Science discourse

Student performance in science discourse varied by dialect condition in each of our three studies. Importantly, we found this result even for the distant peer study, which used a brief and non-interactive paradigm. In other words, we found that students who heard a non-interactive, four minute audio recording of “a student” demonstrated more instances of science discourse from pre-test to post-test, even when the speaker used SAE. The implications of this may be useful for interpreting the results presented here. Learning is largely considered an active process, where increased participation can lead to post-test performance [Wertsch et al., 1993, Chi et al., 1994]. If students’ own science language production was so sensitive to hearing a model that there were differences after such a light intervention, it is likely that students already *had* “*access*” to a sociolinguistic repertoire of science discourse. In this understanding, Jamie’s model may have *elicited* their use of that repertoire, rather than *taught* them that repertoire.

The distant peer study revealed that while there were within-subjects gains in both groups, there were larger gains when the recording was heard in AAE. At post-test, students who heard the recording in AAE demonstrated approximately twice as many instances of science discourse as students who heard recordings from a monodialectal SAE speaker. We are not able to tell *why* students improved after hearing a student model, nor why the differences would be so much greater for the students who heard the model in AAE.

I note here also that the limited number of subjects within the distant peer experiment may have increased the possibility of false positives. Interestingly though, these results happened to “survive” two iterations of annotation. When this work was first carried out, I had a broader definition of what I was referring to as *school-ratified science argumentation* which involved both



discourse and reasoning. As I continued to carry out this work however, I realized that separating out these variables may give additional clarity on what we were seeing. Years after the initial distant peer data was collected, a new team of annotators who had been hired to annotate for our most recent study had *also* re-annotated the distant peer data using the same scheme. We did this to ensure consistency between the three studies. These annotators were blind to condition and to the initial annotations of the distant peer work. The initial result we found with an older annotation scheme remained with our newer conception of our target variable of interest. Using both conceptions of the school-ratified “science discourse” we were aiming to capture, students did *more of it* when they heard the presented model speak using AAE.

We ran two experiments using an interactive virtual peer to analyze if observing students’ interactions with the agent could provide clarity on what the impact of agent dialect might have on students’ experiences in the intervention. I briefly review here the primary outcomes of both studies, and then further aim to provide clarity on what these results might tell us.

In the transfer study, we found a similar result to the distant peer study, even though the science post-test was given the next day. These results play an important role in the work presented in this thesis for two reasons, both of which stem from how this pre/post-assessment was given. First, students’ discourse was collected from an experimenter (a confederate who introduced herself to the students as a teacher from another school). In the other two studies, students presented their science presentations *to Alex* immediately following the intervention. The nature of *that* design means that both immediacy effects as well as effects based on their perception of the character were more likely. For example, in the longitudinal study, we found that rapport during the intervention was predictive of students’ science discourse during the final presentation; as I will describe more in the following section, there may be fairly straight forward and colloquially intuitive interpretations for why that might be.

The second important distinction of how pre/post assessments were given in the transfer study is that they were more structured than the presentation assessments in the distant peer and

longitudinal studies. In those two studies, students were *given the floor* to present their ideas after hearing Alex demonstrate an A+ performance. In this study, the confederate asked the student five questions that each required students to make assessments about whether or not a given environment was a good place for that creature to live.

I emphasize here a point that is important throughout this work, and that I aimed to introduce in the methods section: verbal annotation is a tricky theoretical process. In this work, we faced tension between two theoretical goals in our design of what *counted* as science discourse. The first goal was that we wanted to capture the type of discourse structure that is both often saliently required by the teacher (and the Next Generation Science Standards) that may be associated with something that could have *felt like* “academic talk” [Brown et al., 2017b]. At the same time, we wanted to *set the bar low enough* to see sensitivity between our two conditions. For this reason, we took a very “semantic-neutral” approach to the annotation of students’ presentation of ideas. This was in part due to an intentional theoretical choice, and in part due to difficulty identifying a better way to reasoning “quality.”

To ground the rest of this section, I provide here several examples of students’ answers to the confederate question: *would this be a good place or a bad place for this creature to live, and why?*

### 8.1.1 Student responses annotated for science discourse

**Davon:** No *because I know he does have food but he doesn’t have no water* and everybody littered.

**Lyndon:** Good *because there’s a lot of foods* and strawberries and there’s trees so he could breathe.

**Caden:** Bad *because there’s not water* in a desert, it’s too much shine and some animals like to hibernate. They don’t like to um - they go where some people like to go in the daytime and they like to travel in the night time.

In these examples, Lyndon and Caden both demonstrate an interesting use of science talk that was not captured in our work. Brown et al. [2017b] argues that students are being under-supported in schools because their colloquial or *everyday* explanations for scientific concepts are under-appreciated. He argues for *disaggregate instruction*, which he defines as first letting students explain concepts in a way that is comfortable and familiar to them, and then transitioning them to school-ratified explanations. Brown explicitly recommends letting students refer to carbon dioxide as “the air plants breathe,” without immediately asking them to take on the new vocabulary as well. In the example above, Lyndon voluntarily demonstrates his knowledge that there are environmental benefits to this habitat because the trees contribute oxygen. Caden demonstrates the bridging between everyday and school-ratified science explanations. He uses the scientific term - *hibernate* - and then explains it in terms of a narrative or *everyday* understanding of what the creatures like to do.

In this work, we annotated for an additional feature called *prior knowledge*, which were students’ applications of domain knowledge to the task at hand. In the instances above, Lyndon and Caden’s responses above were both also marked as having prior knowledge. In our initial conception of science discourse, we considered a broader interpretation of discourse to include *vocabulary* (e.g., “hibernate”), *prior knowledge* (e.g., trees being beneficial “so he can breathe”), and *reasoning structure* (e.g., “[claim] because [justification].”

Ultimately, we decided that we did not have a strong theoretical basis to include these different types of speech together as one type of talk. While they would all likely contribute to the broad category of “sounding like a scientist” which motivated this work [?Brown et al., 2005], we decided they should be treated as separate variables to be explored individually, and not as one combined feature.

To prevent over-exploration of a limited sample of data, I focused my analyses on science discourse reasoning structures. There were two reasons for this. First, because we took a semantic-neutral approach to the use of this feature, it would *neutralize against* differences in students’

domain knowledge.

Second, without clear boundaries, it became difficult to interpret what “got to count” as being school-ratified. For example, Davon explains that a given environment is bad because everyone had littered. The picture he was referring to was of a park scene and involved a walking path, trimmed grass, and a bench. There was no visible litter in the picture (frankly, it looked to me like a very well-maintained park!), but it is possible that Davon was applying his knowledge that parks *can become* littered. The park scene was the only one of the four environments that looked readily-accessible to humans, and it is possible Davon was aiming to *get at* the dangers that come from living in a place where people might go. (This was an explanation given by several students in various forms). Similarly, this data was collected in the beginning of May; it is possible Davon recently heard *Earth Day* explanations in school, and that had primed him to apply this understanding to the task.

But it was impossible to know *why* Davon mentioned littering. In working to gain inter-rater reliability with my annotators, this particular example became a salient point of discussion. On first glance, the mention of littering seemed like a throw-away comment, and two annotators wanted to mark it as an example of *non-school ratified evidence*. Another annotator argued that it *could* be school-ratified evidence, and more specifically that it was *also* an example of prior knowledge. The three undergraduate annotators brought up this example in our meeting, and it became the focus of discussion: *what knowledge “counts” as prior knowledge, and how can we draw that line?* Ultimately, for this reason, our science discourse annotations were narrowed to the structures of talk I described in Chapter 3.

### **8.1.2 Student responses not annotated for science discourse**

Though the prior section does emphasize that we aimed to try and *make it easy* for students’ answers to be annotated for science discourse, there were *more* instances of students’ explanations that did not contain the use of these structures. The three examples below demonstrate examples

of the three *types* of non-school ratified science responses students demonstrated.

**Jachelle:** No... Dark... No safe place to swim...too many trees...

**Alexis:** Bad because like fishies - what if he ate one of the fish and then they would know - they would be scared of him and he would - and then he would not want him around and he could like open his mouth and eat these worms and then the worms that saw him would not want to be next to him and tell all the worms to stay away from him, and the flies, well it be good like if not fish were in the water like he could swim around, and go up trees, and the flies - he will get his legs out and eat them and then they will be afraid because they don't want to be eaten.

**Jayla:** Bad place because he will probably get hot. He'll probably look around like "what should I do? What should I eat?" and he don't have to eat I mean drink.

**Breanna:** Bad! Snake! That snake!

Jachelle demonstrates a stream-of-consciousness response which was common among students. These responses often began with an initial claim ("bad" or "good"), and then went on to list a mix of observations, additional claims, and other assertions, with no explicit linking of the ideas together with a school-ratified discourse structure. Alexis demonstrates a *narrative* structure, where she explains the problematic set of circumstances that would arise for a monster if all of his food sources were afraid of him, as well of the resulting negative emotional status of the entire eco-system shown within that environment. This is a compelling explanation from a 2nd grade student about what might happen when a new creature is introduced into an environment. Regardless, its strengths were not captured by our scheme.

Breanna demonstrated a claim and evidence, though without the necessary discourse we were annotating for to link the two ideas. This example demonstrates an admitted reality of our coding scheme. Had Breanna instead said "bad because of that snake!" her response would have been annotated as having science discourse. This example demonstrates an important point present in this work: our annotations of science discourse capture that - discourse. It does not capture

differences in students' domain knowledge, and thus the claims we can make about it was a marker of science performance are limited.

### 8.1.3 Limitations of Wizard of Oz approach

In this work, I found that students were more likely to situate their justifications in science discourse with a bidialectal model in three distinct studies: (1) during an open-ended monologue immediately after hearing a non-interactive example, (2) during a next-day post-test given by a confederate who asked five specific questions, and (3) to a final presentation given to Alex immediately after Alex presented a model during three separate occasions over six weeks.

As I explained, I aimed to focus our science discourse annotations on something we had expected students to be able to demonstrate regardless of prior knowledge. Despite this, there were significant differences by agent dialect in each of these three studies that took place under slightly different conditions. In the distant peer study, the model's content was completely identical for each student, save the dialect in which it was presented. The differences found from this study, and the fact that we internally replicated the result with two slightly different conceptions of science discourse annotated by two separate sets of annotators, lessen the possibility of experimenter bias in this work. However, it is important for us to consider the ways in which this factor could have been introduced during the virtual agent studies.

In explaining the potential relationship between agent dialect and students' science discourse as emerged from our studies, I situate the explanation in the Wizard of Oz approach used in this design. In both of the agent studies, I served as "*the wizard*" who listened in on students' interactions with Alex and selected the speech response Alex gave. In each of the virtual agent studies, there was no significant difference in the number of questions Alex asked, instances of science reasoning Alex demonstrated, nor the amount of discourse Alex demonstrated. Importantly, and perhaps for this same reason, there were also no differences in *students'* use of science reasoning nor science discourse by condition.

This was an initially surprising finding, as we had hypothesized that students who demonstrated more science discourse with Alex would also demonstrate more science discourse at post-test; for this reason, we expected to see significant differences in both students' exchanges to Alex as well as their post-test performance. I posit here two explanations for why students might not have demonstrated differences in science discourse with Alex.

First, Alex took an active guiding role in these dialogues. When students demonstrated an unsupported claim, Alex would ask students to better explain their ideas by demonstrating uncertainty or proposing a potential explanation and asking the student for feedback. Aside from salient examples described in the next section about rapport, this generally served to result in students demonstrating additional reason and discourse.

Second, it is probably that due to the nature of dialogue, students were more easily able to demonstrate science discourse to Alex. Particularly in the peer task that took the form of an informal peer brainstorming task, students would often build off of what Alex was saying (e.g., "yeah and because then he could..."). In these instances, these connectives ("yeah and because") were perhaps easier to access, as they were part of *the repertoire for dialogue* rather than *the repertoire for science*. In other words, using "because" to build off of what someone else was saying may be part of a different *register* than using "because" to independently justify your own science ideas. However, we found no differences in students' use of science discourse during the teacher presentation task, even though these were more structured and required students to more independently build their own ideas.

The strongest limitation introduced by the Wizard of Oz paradigm is in the potential differences introduced when considering rapport. In the following section, I review our findings as they relate to rapport, and then consider what this means when taken together.

## 8.2 Rapport

In the transfer study, we found that there was no significant difference in students' rapport with Alex in the peer task, but there was in the teacher task. We also found that rapport in the teacher task was a significant factor in predicting how much science discourse students demonstrates at post-test in both studies. Though we cannot fully discount other potential factors we did not annotate for in this work, the strongest claim that can be made from both the transfer study and the longitudinal study are that there are differences in students' outward demonstrations of rapport with the agent, and these seem to impact students' subsequent science performance.

### 8.2.1 “That’s what I was thinking too!”

As I stated in the prior section, there were no differences in the number of instances of science discourse Alex demonstrated to students (nor in the amount of science discourse students demonstrated to Alex). However, through our thin slice data annotations, we revealed that Alex's science contributions were not always equally received by students.

In the peer brainstorming task, Alex demonstrated the sorts of vulnerabilities that aligned with learning, such as confessing uncertainty and trailing off sentences. Arguments were also built collaboratively between students and Alex, and Alex would demonstrate a mixture of asking questions (through uncertainty) or proposing new ideas. Among the students who demonstrated high rapport throughout the session with Alex (and who ultimately did better at post-test), students' outward affective displays (whether positive or negative) were often triggered by a response that Alex gave.

It was not rare for students to demonstrate bursts of outwardly positive-seeming affect in response to Alex's contributions, though it is not always clear under what circumstances these behaviors are elicited. The example below is from a 2nd grade student in the transfer study who was working with bidialectal Alex.

**Alex:** Why you think he got those spike on his back?



**Caden:** Those sp- those long, like... they look all...

**Alex:** Those spike on his back look like, almost like a porcupine, but not as sharp.

**Caden:** Yeah! And so if someone is trying to run into them, then he can protect?

**Alex** (*hesitant*): Wait... why that?

**Caden:** Because, you see how he - okay you see how he's not very sharp in the teeth?

**Alex:** Uh huh?

**Caden:** If he don't have sharp teeth then he can't defend.

**Alex:** Oh, yeah! That make sense. I think he can use those spike to protect himself from animals running into him because he don't got a lot of other sharp ways to protect himself.

**Caden** (*visibly excited*): Yeah! That's what I think too!

In this exchange, Alex demonstrates one of the pedagogical principles that has been often cited in pedagogical agent literature. Because students generally learn more from *explaining* than *receiving explanations*, Alex could elicit student explanations by asking for explanations in ways that sounded like the explanation was needed because she was unsure, rather than because she was quizzing the student.

### 8.2.2 “How are you this dumb?”

In contrast to the example above, students' outward responses to these sorts of behaviors from Alex were not always positive. In fact, from the content of how students responded to Alex, it seems that the agent's good intentions to *ask for more information* were not perceived as a genuine desire to learn from the student, but perhaps as Alex ignoring the student's ideas. The example below was also taken from the transfer study, and involved a student working with the monodialectal agent.

**Alex:** I wonder, why do you think he has those horns - um, antlers?

**Saniya *starting a new topic*:** Alex. Um, he probably don't get a lot of water from right there, from right here, on the picture, because there's no water

**Alex:** I wonder how he drinks if there isn't any water around

**Saniya:** He probably goes to a different land and drinks the water there.

**Alex:** Yeah. Hm. What if... I wonder, what if he can drink from the trees? Like when it rains?

**Saniya:** Yeah.

**Alex:** I'm not sure about it though... what do you think?

**Saniya:** Like I was *saying*, another picture.

Several things are demonstrated in this exchange. As I illustrated in the general dialogue flow diagram presented in Chapter 5, Alex would guide students towards eliciting reasoning by asking specific questions or demonstrating uncertainty. In this instance, the student proposed an unsupported claim: *he probably goes to a different land*. In return, Alex provided a prompt that could guide the student toward additional support for her claim, though she did not pick up the bait ("yeah."). Alex then demonstrated uncertainty and asked the student her perspective. This sort of verbal move was generally a productive way to elicit students' responses, as in the example shown in their prior section. In this instance however, the student seemed to demonstrate frustration at Alex's question, perhaps feeling as if Alex had been ignoring her.

Students also occasionally interpreted Alex's questions as lack of intelligence. In the longitudinal study chapter, I provide a student case study of one particular student, Terran, who seemed to view Alex's questions or contributions as lack of intelligence. Terran was an interesting case study for interpreting the relationship between rapport and science discourse use. Especially during the later sessions in the longitudinal study, beginning around the third session, Terran began spending his time with Alex demonstrating behaviors that appeared to be instances of *breaking* the agent that I annotated as *social challenges*, but which are more widely referred to as *agent abuse*. I describe this concept further below.

## 8.3 Theoretical perspectives

### 8.3.1 Increased practice

From a cognitive perspective on peer tutoring, dialogic learning environments evoke opportunities for students to organize their own reasoning and recognize misconceptions through their ability to explain their ideas to another person. This idea has been corroborated through the evidence that students are more likely to learn from *giving* explanations than *receiving* them [Webb, 1989, VanLehn et al., 2007, King, 1992]. In the transfer study, our data supports these results. When controlling for pre-test, a model that includes both *condition* and *science discourse during the teacher presentation task* finds that it is increased science discourse with Alex, and not condition directly, that improves students post-test performance. However, it appears that condition plays an important role in determining how much science discourse is presented to Alex during the teacher task, even though there aren't differences in this behavior directly. Specifically, our results demonstrate that increased rapport with Alex is associated with increased rapport during the teacher task; in turn, increased rapport in the teacher task is associated with post-test science discourse.

It remains to be understood however why increased rapport would lead to increased levels of science discourse. Below, I consider two potential perspectives that have been argued to have relevance from bidialectal pedagogy literature: second language acquisition and second language participation. In the case of the studies presented here, we investigated two types of language styles within a larger register referred to as *academic language*. In second language literature, this distinction broadly refers to *acquisition* as learning a new way of speaking as a set of individual nouns, grammatical properties, and so forth (e.g., a lesson where one learns how to translate between two different language styles, or is studying new vocabulary terms). Second language participation on the other hand refers to *picking up* new language styles more holistically, through gaining familiarity with a certain *way of speaking* and then choosing to

access that sociolinguistic repertoire in intentional ways.

This distinction has been made using fairly consistent terminology in the linguistics literature [Siegel, 2006a, 2008, Bialystok, 2009]. It has *also* been described in the *science language* literature. Scholars like Lemke [?] and Brown [Brown et al., 2017b] have argued that *science is a second language*, and we should be treating its learning as such.

Though there is little work talking about learning both an *English dialect style* and a *science discourse style* together, there is some work that is important and associated. Okhee Lee [Lee and Fradd, 1998] argues that it is important in the education of language minority students, it may be *particularly* hard for them to additionally learn how to demonstrate the use of school-ratified science discourse.

Carol Lee [Lee et al., 2005] proposes that instructional design that both integrates and allows students' home language styles in the classroom may have substantial pedagogical benefits for under-served students. Though she does not draw a clear distinction in this call between whether these systems might help students *acquire* new language styles or rather *participate* in new language styles, she argues from a perspective of praxis: students do better in environments that make use of more culturally familiar language behaviors.

### 8.3.2 Second language acquisition

From a *second language acquisition* perspective, bidialectal pedagogies may work because they promote students' comprehension of instruction [Nieto, 1999], such as through reducing the cognitive load of hearing information in a way that is less immediately easy to interpret [Sweller, 1988]. Edwards [Edwards et al., 2010] argues that each student may be at distinct stage of acquiring fluent bidialectal proficiency. Prior to achieving fluency in two language styles, code switching may entail needing to consciously override a dominant response, which may take substantial cognitive effort [Bettenhausen and Murnighan, 1991, Feldman, 1984, Bialystok, 2009].

From a cognitive perspective, it is possible that the reason students succeeded with the bidi-

alectal agent was due to increased ease of understanding, and thus a reduced cognitive load. Though we can not fully rule out this interpretation, our results do not actively adduce evidence for this explanation. Terry and Smith [Terry et al., 2010] posit that one way of assessing whether or not students are being limited from cognitive factors within an intervention is to identify whether there is a linear relationship between how much dialect they demonstrate and their resulting performance. This is of course a simplified explanation of what is surely a complex phenomenon; regardless, the results from either of the two agent studies (nor the initial study using the non-interactive “peer recording”) adduce any evidence for the possibility that high-dialect speakers fared worse with the monodialectal agent than low-dialect speakers.

### 8.3.3 Second language participation

Scholars of socio-cultural perspectives highlight the historical contexts, attitudes, and identity orientations that are embedded into all interaction, including learning [Mason, 2007]. Within these perspectives, the metaphor driving learning is not acquisition but *participation* Mason [2007]. Brown (Brown et al. [2017b], Brown [2004]) argues that educators should not presuppose that all students can take on the language requirements of the classroom free from social complication. From this perspective, language has the power to indicate social group affiliation and membership, even within (or, from some lenses, at the expense of) educational environments [Fishman, 1989]. For example, Godley’s ([?]) ethnography on high school students revealed that students reported they *could* speak SAE, but that they reserve its use for *important things* like job interviews, not just classroom interactions. Similarly, Ochs ([Ochs, 1993]) identified that some students associate AAE with their *disidentification* from an academic identity.

#### Cultural salience

Evidence suggests negatively stereotyping computers once they are endowed with a voice [Nass et al., 1996] or a face [Sproull et al., 1996]. With a few exceptions [De Angeli and Carpenter,

2005, De Angeli et al., 2006], research has ignored potential negative outcomes of the interaction, despite the fact social agency subsumes the occurrence of conflicts. Even in cases where purely descriptive research is conducted, negative user responses to virtual partners are normally ignored.

Scholars have argued that some marginalized students may not want to participate in an education system that not only denigrates their language and culture, but also privileges students from upper middle class social groups [Giroux, 1988]. Giroux has argued that this is one potential reason that students may demonstrate a tension when school environments promote an exclusively SAE environment. There is a concerning tension that emerges in some dialogues about the under-performance of African American students where it is spoken as if it is *a problem with culture*, but it is critical that assessments such as those are not taken out of the context in which they exist. To be academically successful in school often means *leaving out* parts of yourself from discussion in a way that not all students seem to have to do; some scholars argue that the tension between *the nature of school* and students' perceptions that they are not valued within the classroom that can result in a disidentification from academic tasks [Siegel, 2006b, Fordham and Ogbu, 1986, Lanehart, 1998].

Though I did not describe this in depth above, our data reveal that students seemed to find some aspects of Alex's behaviors culturally salient. Though it was not overwhelmingly frequent, students did make references to Alex's perceived race and occasionally dialect. Some of these comments were based in familiarity, and were captured by our annotations of positivity. For example, one student shared, "Oh I like that hairdo Alex, I like wearing mine like that sometimes too." More frequently, references to Alex's visual appearance or language behaviors occurred as social challenges. Interestingly, 100% of these cases occurred in the monodialectal condition. In the fourth session of the six week longitudinal experiment, Terran (the student described in the longitudinal case study), began referring to Alex as *slave boy* whenever Terran was in the role of teacher.

Fordham [Fordham, 1998] finds that African American students who opt for academic success often develop a strategy of racelessness that is, avoiding being identified with any particular racial group. In turn however, they may be seen by others as being *cultural traitors*. Scholars have argued that when students who have begun to *disidentify* with school, there may be benefits to supporting the development of positive, integrated, and complementary attitudes towards their cultural participations in both their social groups and their school [Boykin, 1994, Seiler, 2001, Boykin and Noguera, 2011]. African American students demonstrate improved academic performance when they feel a sense of community within their learning environment [Robinson and Biran, 2006], and feelings of community, collaboration, and kinship have been shown to have greater impacts on the positive performance of these students than their white peers [Carson, 2009]. It is possible that these social mechanisms may be at play in helping to interpret why Alex's dialect seemed to have impacts on the social behaviors of these students.

It is possible that through doing so, Alex was able to represent the difference between *a cultural other* or *a member of their community*, and that this perception impacted their own resulting science behaviors. Though we cannot say for certain, our results regarding students' dialect attitudes may additionally support this interpretation. The results of our transfer study demonstrated that students who worked with the monodialectal agent demonstrated *less positive attitudes about SAE*. Our analyses revealed that the more students demonstrated social challenges to Alex, the more likely they were to feel more negatively about SAE at post-test.

## **Signifying**

Importantly, a sociocultural lens reveals that this may be a *particularly* possible interpretation among African American youth [Heath, 1983, Corsaro and Eder, 1990]. Ethnographers who have studied communication norms among African American youth have found that clever and elaborated demonstrations of social challenges are one way of *building* social bonds. These communication norms have been referred to as *playing the dozens* or *signifying*, and have been identified

among linguists for their profound verbal playfulness and sophistication. Heath [Heath, 1989] describes that these behaviors are so common among children that even toddlers have learned that one way to avoid punishment for wrong-doing is to impress adults with verbal demonstrations that can distract from the unpleasant interaction.

The take-away from these above explanations that I emphasize again here that there may be complex explanations for why these social challenges, perhaps in some instances *signifying behaviors*, took place. Importantly, it may be the case that Alex was unable to demonstrate in-group behaviors through the use of AAE, and students were investigating if Alex were able to *instead* demonstrate in-group behaviors through the use of mutual social challenges. Corsaro [Corsaro, 1992] describes these behaviors in her ethnographies on pre-school African American and White American students. In her work, she finds that *argumentation* among White students are often in situations that seem personally stressful, while *argumentation* among Black students were more common in situations that felt like play. It may therefore be that what *looked* like outward displays of resistance may hold different sociocultural meaning among the students.

In prior work I completed outside the scope of this thesis with collaboration from two members of this committee, Amy Ogan and Justine Cassell, we identified that *teasing* played a role in the learning gains of pairs of (human) students engaging in a peer tutoring task [Wang et al., 2012]. In this study, we defined teasing as utterances that had observable aspects of both positivity and negativity (e.g., "*love ya, jerkface*" or "*good job... finally*"). However, we also found that the *type* of impact teasing had was moderated by the students' reported friend status. In that study (not controlled within the experimental design), students either participated with a self-identified friend, or were assigned a stranger or acquaintance to work with who was not an identified friend. Among friends, teasing was positively predictive of learning gains; among strangers, teasing was *negatively* predictive of learning gains.

I bring up this study here to emphasize an important point: social challenges may not be inherently negative in all cases. In fact, as in the "*love ya, jerkface*" example above, there



are some instances where certain types of social challenges are not aggressive, but rather are demonstrations of closeness. This is in line with theories of social bonding referred to *violations of social norms*, wherein relationships among strangers are expected to closely follow one's best understanding of social norms for a given context, and then *violation* of those norms can increase over time to mark affection. In the study described above, we offered the interpretation that students' teasing was one way of *evening out the playing field* in a socially threatening academic task. Among strangers, however, there may not be the same *social safety net* afforded by pre-existing affection to have the same positive role. Without the pre-existing bond of friendship, such teasing may instead be seen as a socially aggressive offense, rather than a way to safely mark affection.

In the present study, we did identify that social challenges presented by the student *was* associated with lower rapport. However, such a finding may be a manifestation of Alex's inability to successfully engage in much sophisticated *reciprocal* social challenging. It may have been the case that students with the monodialectal agent did in fact feel the increased academic pressure of the teacher task, and turned to *norm violations* or *social challenges* as a way to level the power between both interlocutors. This phenomenon has been described by some scholars as the demonstration of *resistance behaviors*, as efforts to mark outward disidentification from academic environments that did not feel in line with personal values.

Regardless of the reasons behind *why* students demonstrated this reduction of rapport or increase of social challenges, it seems that they do play an important role on students' learning outcomes. These behaviors were associated with less school-ratified science behaviors demonstrated towards Alex during the intervention, and less school-ratified science behaviors at post-test. In fact, students who demonstrated more social challenges to Alex during the teacher task were even significantly more likely to report a personal reduction in their own attitudes about Standard English at post-test.

## **(Dis)identification**

Due to its prevalence in our dataset and the seeming importance of its implications for our results, it would be remiss to not discuss agent abuse. It is likely that instances of agent abuse are occurring more frequently in pedagogical data sets than are being reported [Zdenek, 2006]. This may be due to embarrassment about the ways in which systems failed to work, or perhaps more likely a consideration of such exchanges as noise in the data set. Regardless, there is evidence that these behaviors are common in interactions with agents. For this reason, De Angeli and Brahnham 2008 [De Angeli and Brahnham, 2008] has questioned the often-cited assertion made by Reeves and Nass that computers treat humans like people [Nass et al., 2000].

While that is certainly true in some ways that I overviewed in the background chapter of this work, scholars who have studied agent abuse have identified the ways in which it *breaks social norms* in ways that are not widely seen among interactions with another human whom you’ve just met. Wallis [Wallis, 2005] takes as a premise that testing an agent’s abilities is a demonstration of a user trying to *break* the agent. Veletsianos [Veletsianos et al., 2008] also refers to this concept of breaking, *or testing*, as ways of demonstrating frustration, disengagement, or boredom while talking with a virtual charter.

On the other hand, in some instances children *do* demonstrate frustration, disengagement, or boredom while talking to human peers as well. Bullying is associated with dominance status and peer leadership [Pellegrini, 1995]. According to Wright and colleagues, prosocial behavior is consistently related to peer acceptance across groups, whereas negative behaviors such as aggression are only associated with peer rejection when a person does not fit in with (i.e., deviates from) other members of the group. Thus, the group norm, meaning the behavior on a group level, moderates the relationship between individual behavior and social status.

With virtual agents, De Angeli and Carpenter [De Angeli and Carpenter, 2005] have identified that users may try to *gain power* in dialogue with an agent by trying to break it, as a way of *putting the computer “in its place.”* In some perspectives on what it means to dialogue with a

virtual agent however, these are signs that the characters are taking on a social meaning to users [Brahnam and De Angeli, 2008]. Brahnam 2009 [Brahnam, 2009] proposes that there is an *artistic need* to create agents that seem *credible*, so as to not reproduce essentially the idea described by the uncanny valley effect: that agents that demonstrate *some* social behavior, but not *enough* or *right* social behavior, may be prime targets of abuse from users. With this understanding in mind, there are at least two broad interpretations of students' increased use of social challenges towards a monodialectal version of Alex, and their resulting reduced science discourse in the intervention that limited their post test learning.

In one interpretation, Alex successfully *embodied* the identity of either *an African American child who only speaks using Standard English, even when there isn't anyone around to tell him not to* or *an American American child who uses non-standardized dialect features (like the students are) while doing the instructed task*. In other words, it might not be the *dialect alone* that resulted in students' directed challenges at the agent, but rather what *sort of student* Alex represented. Wright et al. [Wright et al., 1986] have described a person-group dissimilarity model that can explain the varying associations between social status and aggression. These theories state that similarity in attitudes does not necessarily lead to attraction, but that dissimilarity in attitudes, on the other hand, is a strong predictor of repulsion [Rosenbaum, 1986, Singh and Ho, 2000].

Though we limit our interpretations here, our secondary analyses within the transfer study seem to indicate that the more students demonstrated *social challenges* to Alex, the more likely they were to reduce their perceptions of Standard English at post-test. Conversely, for students who worked with the bidialectal agent, there was an inverse relationship between students' attitudes about Standard English at pre-test and how much science discourse they demonstrated with the agent. In other words, the students who were *most* successful with Alex were the ones who worked with the bidialectal agent and reported negative attitudes at pre-test about the dialect style most commonly associated with school.



## Chapter 9

### Concluding remarks

*"While the media and public discourse attacked Black Language (BL) and Black people for so-called "deficiencies," a generation of young Hip Hop Headz (including me) spent hours crafting linguistic skillz and pushin the boundaries of the English language in rhyme ciphers, battles, and freestyles. Wasn't no way in the world you could get me to see BL as deficient!" - Samy Alim, Director of Stanford's Center for Race, Ethnicity, and Language. (Alim et al. [2008], pp. 24).*

In this work, I investigate the effectiveness of an interactive technology called a virtual peer that models the use of science discourse during dialogues with language minority African American students. In the work presented here, I bridge theories from sociolinguistics and cultural theory with the possibilities introduced through educational technologies. In doing so, I investigate how cultural biases or unquestioned assumptions within the design of an educational technology may ultimately impact the effectiveness of that system for some populations. As improving the educational outcomes for African American students remains a national priority, the work presented in this thesis focuses on evaluating the impact of one theoretical approach that has been long-argued to be able to support certain language minority populations: dialect integrated instruction. There is well-supported but controversial evidence that African American students improve in classrooms that take a culturally-aligned approach to pedagogy, such as by integrat-

ing student' home dialects into the learning environment. At the same time, the dialect used by educational technology systems is largely an unconsidered design decision. By default, educational technologies, traditional learning environments, and much of society at large considers many *facts of reality* to be culturally neutral; this question sits at the center of this investigation.

In this thesis, I demonstrate that the dialect in which a pedagogical agent models science discourse has impacts on students' social behaviors during the interaction, their science discourse with the agent, and ultimately their science discourse at post-test. The relationships between these variables however make open up more questions than they answer. Though these studies cannot make conclusive contributions to these complex bodies of theoretical work, they adduce evidence for the arguments which posit benefits of bidialectal instructional environments for language minority students. This work is presented here in an intersection of tension between potential pedagogical benefits and the folk theories of well-meaning educators who have warned me, "*students can't learn anything in broken English.*" In terms of both the assertion of African American English as *broken English* as well as its potential for supporting student achievement, our results add to an ever growing body of literature demonstrating that these folk theories are just not the case.

Further, as our results demonstrate these pedagogical benefits within an educational technology, these studies are the first to my knowledge to evaluate how the theories in support of dialect integrated learning environments may apply to new technological systems. These results adduce evidence for the arguments that technological systems carry cultural weight. While we typically view technologies as deracialized artifacts, I aim for this study to serve as an additional reminder that there aren't culturally neutral choices at all - particularly when it comes to education [Henderson, 2013].

For these same reasons, I also use this thesis as a call for instructional designers of both real and virtual learning environments to consider what it means to design a culturally equitable learning environment. Often, in a well-meaning desire to promote a positive environment

for marginalized students, designers of educational technologies may identify and manipulate surface-level characteristics (e.g., skin color) [Moreno and Flowerday, 2006] or stereotyped artifacts (e.g., hair braiding) [Eglash et al., 2013] to promote student engagement without *also* situating these culturally marked behaviors within a culturally competent context [Henderson, 2013].

In the work presented here, an African American virtual peer who only spoke Standard English received a categorically different response from students in terms of received social challenges than a character who demonstrated in-group use of the dialect system used by all students in our studies. Without the use of this dialect, Alex was subjected to far greater ridicule, testing, insults, and other social challenges that are prohibitive to a productive learning environment. As we found that these results maintained over a six week intervention, novelty effects may not be a wholly sufficient explanation for these social behavior differences. When these findings are analyzed through the same bodies of literature on which these studies were designed, it may be the case that agent abuse is not simply *noise in the data*, but evidence that we as designers of educational technologies are failing to meet a social need.

These findings contribute to an growing body of evidence that that virtual agents that are designed with *some* but not *enough* or *right* social cues may have negative impacts for the students who we are trying to support. As [Moreno and Flowerday, 2006] found in her own work, students of color who chose to work with a “same ethnicity” virtual agent did not learn from the intervention as much as students who either chose or were assigned to work with a white agent. In reviewing her results, Moreno et al. posit that students may have been frustrated by or curious about the lack of cultural alignment demonstrated in the agent, which may have *distracted them* from the task at hand. This paper ends with a warning that it may be too difficult to *design for culture appropriately*, and that instead it may be important to *remove unnecessary social cues from the learning environment*.

I end this thesis by positing that we cannot remove unnecessary social cues from a learning

environment. At best, the *lack* of inclusion of social cues *is a social cue in and of itself*. There is *ideological baggage* tied up not just in the *way* we educate students, but at its very core *what* we teach students, and what *version of themselves* we are trying to build through their education. Alexander (2007, p. 67) argues that efforts to fully distinguish between whether a mechanism driving students' learning is based in *cognitive mechanisms* or *social mechanisms* are either unnecessary or unachievable. As with most frameworks, neither alone accounts for the inherent complexities within our broad conception of education. Lave and Wenger (p. 53) use this understanding to explain the ways in which *classrooms are communities*, proposing that “identify, knowing, and social membership entail one another” [Lave and Wenger, 1998]. In the case of educational technology design, it is important for us to take formal notice about whose culture is being recognized, privileged, validated, and lifted up through our choices. Students' emotional, social, personal, and political contexts are a reality from which they cannot be removed; for this reason, neither can be our approaches towards their education.



# Bibliography

- Jessica Abrams, Joan OConnor, and Howard Giles. Identity and intergroup communication. *Handbook of international and intercultural communication*, 2:225–240, 2002.
- Carolyn Temple Adger and Donna Christian. Sociolinguistic variation and education. *Sociolinguistic variation: Theories, methods, and applications*, 2007.
- H Samy Alim. *Roc the mic right: The language of hip hop culture*. Routledge, 2006.
- H Samy Alim. Hip hop nation language. *Linguistic anthropology: A reader*, pages 272–289, 2009.
- H Samy Alim, Awad Ibrahim, and Alastair Pennycook. *Global linguistic flows: Hip hop cultures, youth identities, and the politics of language*. Routledge, 2008.
- Elliot Aronson and J Merrill Carlsmith. Experimentation in social psychology. *The handbook of social psychology*, 2(2):1–79, 1968.
- Joshua Aronson, Carrie B Fried, and Catherine Good. Reducing the effects of stereotype threat on african american college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, 38(2):113–125, 2002.
- Doris Ash. Reflective scientific sense-making dialogue in two languages: The science in the dialogue and the dialogue in the science. *Science Education*, 88(6):855–884, 2004.
- M Astrid, Nicole C Krämer, Jonathan Gratch, and Sin-Hwa Kang. it doesnt matter what you are! explaining social effects of agents and avatars. *Computers in Human Behavior*, 26(6):

1641–1650, 2010.

Robert K Atkinson. Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*, 94(2):416, 2002.

Robert K Atkinson, Sharon J Derry, Alexander Renkl, and Donald Wortham. Learning from examples: Instructional principles from the worked examples research. *Review of educational research*, 70(2):181–214, 2000.

Robert K Atkinson, Richard E Mayer, and Mary Margaret Merrill. Fostering social agency in multimedia learning: Examining the impact of an animated agents voice. *Contemporary Educational Psychology*, 30(1):117–139, 2005.

Kathryn Au, Cathie Jordan, et al. Teaching reading to hawaiian children: Finding a culturally appropriate solution. *Culture and the bilingual classroom: Studies in classroom ethnography*, pages 139–152, 1981.

Jeremy N Bailenson, Jim Blascovich, Andrew C Beall, and Jack M Loomis. Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, 29(7): 819–833, 2003.

Cynthia Ballenger. Social identities, moral narratives, scientific argumentation: Science talk in a bilingual classroom. *Language and Education*, 11(1):1–14, 1997.

Albert Bandura. Principles of behavior modification. 1969.

Albert Bandura and Daniel Cervone. Differential engagement of self-reactive influences in cognitive motivation. *Organizational behavior and human decision processes*, 38(1):92–113, 1986.

John A Bargh and Yaacov Schul. On the cognitive benefits of teaching. *Journal of Educational Psychology*, 72(5):593, 1980.

S Barlow, S Converse, S Kahler, J Lester, and B Stone. Animated pedagogical agents and problem solving effectiveness: A large scale empirical evaluation. In *Artificial intelligence in*

*education*, page 23, 1997.

Reuben M Baron and David A Kenny. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173, 1986.

Angela Calabrese Barton and Edna Tan. Funds of knowledge and discourses and hybrid space. *Journal of Research in Science Teaching*, 46(1):50–73, 2009.

Angela Calabrese Barton and Edna Tan. We be burnin’! agency, identity, and science learning. *The Journal of the Learning Sciences*, 19(2):187–229, 2010.

Amy Baylor and Yanghee Kim. Validating pedagogical agent roles: Expert, motivator, and mentor. In *EdMedia: World Conference on Educational Media and Technology*, pages 463–466. Association for the Advancement of Computing in Education (AACE), 2003.

Amy L Baylor and Soyoung Kim. Designing nonverbal communication for pedagogical agents: When less is more. *Computers in Human Behavior*, 25(2):450–457, 2009.

Amy L Baylor and Yanghee Kim. Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. In *Intelligent tutoring systems*, volume 3220, pages 592–603. Springer, 2004.

Amy L Baylor and Yanghee Kim. Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education*, 15(2):95–115, 2005.

Charles Bazerman et al. *Shaping written knowledge: The genre and activity of the experimental article in science*, volume 356. University of Wisconsin Press Madison, 1988.

Carl Bereiter. Implications of postmodernism for science, or, science as progressive discourse. *Educational Psychologist*, 29(1):3–12, 1994.

Thomas J Berndt. Friends’ influence on students’ adjustment to school. *Educational Psychologist*, 34(1):15–28, 1999.

Thomas J Berndt and Keunho Keefe. Friends’ influence on school adjustment: A motivational

- analysis. 1996.
- Basil Bernstein. Language and social class. *The British journal of sociology*, 11(3):271–276, 1960.
- Kenneth L Bettenhausen and J Keith Murnighan. The development of an intragroup norm and the effects of interpersonal and structural challenges. *Administrative Science Quarterly*, pages 20–35, 1991.
- Ellen Bialystok. Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and cognition*, 12(1):3–11, 2009.
- Ulf Bichel. *Problem und Begriff der Umgangssprache in der germanistischen Forschung*, volume 32. Niemeyer, 1973.
- Timothy Bickmore, Amanda Gruber, and Rosalind Picard. Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient education and counseling*, 59(1):21–30, 2005.
- Gautam Biswas, Krittaya Leelawong, Daniel Schwartz, Nancy Vye, and The Teachable Agents Group at Vanderbilt. Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, 19(3-4):363–392, 2005.
- Monique Boekaerts. Understanding students affective processes in the classroom. *Emotion in education*, 1:37–56, 2007.
- A Wade Boykin. Afrocultural expression and its implications for schooling. *Teaching diverse populations: Formulating a knowledge base*, 1944:50–51, 1994.
- A Wade Boykin and Pedro Noguera. *Creating the opportunity to learn: Moving from research to practice to close the achievement gap*. Ascd, 2011.
- BA Bracken. *Bracken Basic Concept Scale*. The Psychological Corporation, San Antonio, TX, 2984.
- Sheryl Brahnam. Building character for artificial conversational agents: Ethos, ethics, believ-

- ability, and credibility. *PsychNology Journal*, 7(1), 2009.
- Sheryl Brahnam and Antonella De Angeli. Special issue on the abuse and misuse of social agents, 2008.
- Holly P Branigan, Martin J Pickering, Jamie Pearson, and Janet F McLean. Linguistic alignment between people and computers. *Journal of Pragmatics*, 42(9):2355–2368, 2010.
- Susan E Brennan. The grounding problem in conversations with and through computers. *Social and cognitive approaches to interpersonal communication*, pages 201–225, 1998.
- Bryan A Brown. Discursive identity: Assimilation into the culture of science and its implications for minority students. *Journal of Research in Science Teaching*, 41(8):810–834, 2004.
- Bryan A Brown. it isn't no slang that can be said about this stuff: Language, identity, and appropriating science discourse. *Journal of Research in Science Teaching*, 43(1):96–126, 2006.
- Bryan A. Brown and Eliza Spang. Double talk: Synthesizing everyday and science language in the classroom. *Science Education*, 92(4):708–732, 7 2008. ISSN 00368326. doi: 10.1002/sce.20251. URL <http://doi.wiley.com/10.1002/sce.20251>.
- Bryan A. Brown, John M. Reveles, and Gregory J. Kelly. Scientific literacy and discursive identity: A theoretical framework for understanding science learning. *Science Education*, 89(5):779–802, 2005. doi: 10.1002/sce.20069.
- Bryan A Brown, Charmaine Mangram, Kathy Sun, Keith Cross, and Erin Raab. Representing racial identity: Identity, race, the construction of the african american stem students. *Urban Education*, 52(2):170–206, 2017a.
- Bryan A Brown, Kihyun Ryoo, Jamie Rodriguez, Bryan A Brown, Kihyun Ryoo, and Jamie Rodriguez. Pathway Towards Fluency : Using disaggregate instruction to promote science literacy Pathway Towards Fluency : Using disaggregate instruction to promote science literacy. 0693(June), 2017b. doi: 10.1080/09500690903117921.
- Donn Erwin Byrne. *The attraction paradigm*, volume 11. Academic Pr, 1971.

- Sandra L Calvert, Melissa N Richards, A Jordon, and D Romer. Childrens parasocial relationships. *Media and the well-being of children and adolescents*, pages 187–200, 2014.
- Rafael A Calvo and Sidney K D’Mello. *New perspectives on affect and learning technologies*, volume 3. Springer Science & Business Media, 2011.
- Leslie R Carson. i am because we are: collectivism as a foundational characteristic of african american college student identity and academic achievement. *Social Psychology of Education*, 12(3):327–344, 2009.
- Justine Cassell, Tom Stocky, Tim Bickmore, Yang Gao, Yukiko Nakano, Kimiko Ryokai, Dona Tversky, Catherine Vaucelle, and Hannes Vilhjálmsón. Mack: Media lab autonomous conversational kiosk. In *Proc. of Imagina*, volume 2, pages 12–15, 2002.
- Courtney B Cazden. The language of teaching and learning. *The language of teaching and learning*, pages 348–369, 2001.
- Carol KK Chan, PJ Burtis, Marlene Scardamalia, and Carl Bereiter. Constructive activity in learning from text. *American educational research journal*, 29(1):97–118, 1992.
- Anne H. Charity, Hollis S. Scarborough, and Darion M. Griffin. Familiarity with School English in African American children and its relation to early reading achievement. *Child Development*, 75(5):1340–1356, 2004a. ISSN 00093920. doi: 10.1111/j.1467-8624.2004.00744.x. URL <http://www.ncbi.nlm.nih.gov/pubmed/15369518>.
- Anne H Charity, Hollis S Scarborough, and Darion M Griffin. Familiarity with school english in african american children and its relation to early reading achievement. *Child development*, 75(5):1340–1356, 2004b.
- Emna Cherif and Jean-François Lemoine. Human vs. synthetic recommendation agents voice: The effects on consumer reactions. In *Marketing at the Confluence between Entertainment and Analytics*, pages 301–310. Springer, 2017.
- Michelene TH Chi. Quantifying qualitative analyses of verbal data: A practical guide. *The*

*journal of the learning sciences*, 6(3):271–315, 1997.

Michelene TH Chi and Kurt A VanLehn. The content of physics self-explanations. *The Journal of the Learning Sciences*, 1(1):69–105, 1991.

Michelene TH Chi, Miriam Bassok, Matthew W Lewis, Peter Reimann, and Robert Glaser. Self-explanations: How students study and use examples in learning to solve problems. *Cognitive science*, 13(2):145–182, 1989.

Michelene TH Chi, Nicholas Leeuw, Mei-Hung Chiu, and Christian LaVancher. Eliciting self-explanations improves understanding. *Cognitive science*, 18(3):439–477, 1994.

James S Chisholm and Amanda Godley. Dialect variation, identity and power: A discourse analysis of bidialectal high school students small group discussion. In *Conference of the International Society for Language Studies, Orlando, FL*, 2009.

Chih-Yueh Chou, Tak-Wai Chan, and Chi-Jen Lin. Redefining the learning companion: the past, present, and future of educational agents. *Computers & Education*, 40(3): 255–269, 4 2003. ISSN 03601315. doi: 10.1016/S0360-1315(02)00130-6. URL <http://linkinghub.elsevier.com/retrieve/pii/S0360131502001306>.

Antonius HN Cillessen and Amanda J Rose. Understanding popularity in the peer system. *Current Directions in Psychological Science*, 14(2):102–105, 2005.

Geraldine Clarebout and Jan Elen. In search of pedagogical agents’ modality and dialogue effects in open learning environments. *E-Journal of Instructional Science and Technology*, 10(1):n1, 2007.

Geraldine Clarebout, Jan Elen, W Lewis Johnson, and Erin Shaw. Animated pedagogical agents: An opportunity to be grasped? *Journal of Educational multimedia and hypermedia*, 11(3): 267–286, 2002.

Richard E Clark and Sunhee Choi. Five design principles for experiments on the effects of animated pedagogical agents. *Journal of Educational Computing Research*, 32(3):209–225,

2005.

Dennis J Condrón. Social class, school and non-school environments, and black/white inequalities in children's learning. *American Sociological Review*, 74(5):685–708, 2009.

William A Corsaro. Interpretive reproduction in children's peer cultures. *Social Psychology Quarterly*, pages 160–177, 1992.

William A Corsaro and Donna Eder. Children's peer cultures. *Annual review of sociology*, 16(1):197–220, 1990.

National Research Council et al. *America becoming: Racial trends and their consequences*, volume 1. National Academies Press, 2001.

HK Craig, L Zhang, SL Hensel, and EJ Quinn. African American English-speaking students: An examination of the relationship between dialect shifting and reading outcomes. *Journal of Speech, Language and ...*, 52(August):1992–2007, 2009a. ISSN 1092-4388. doi: 10.1044/1092-4388(2009/08-0056). URL <http://jslhr.asha.org/cgi/content/abstract/52/4/839>.

Holly K Craig. Effective language instruction for african american children. 2008.

Holly K Craig and Julie A Washington. Grade-related changes in the production of african american english. *Journal of Speech, Language, and Hearing Research*, 47(2):450–463, 2004.

Holly K Craig and Julie A Washington. *Malik goes to school: Examining the language skills of African American students from preschool - 5th grade*. Psychology Press, 2006.

Holly K Craig, Julie A Washington, and Connie Thompson-Porter. Average c-unit lengths in the discourse of african american children from low-income, urban homes. *Journal of Speech, Language, and Hearing Research*, 41(2):433–444, 1998.

Holly K Craig, Lingling Zhang, Stephanie L Hensel, and Erin J Quinn. African american english-speaking students: An examination of the relationship between dialect shifting and reading outcomes. *Journal of Speech, Language, and Hearing Research*, 52(4):839–855,



2009b.

Scotty D Craig, Barry Gholson, Joshua K Brittingham, Joah L Williams, and Keith T Shubeck.

Promoting vicarious learning of physics using deep questions with explanations. *Computers & Education*, 58(4):1042–1048, 2012.

Scotty D Craig, Jessica Twyford, Norma Irigoyen, and Sarah A Zipp. A test of spatial contiguity for virtual humans gestures in multimedia learning environments. *Journal of Educational Computing Research*, 53(1):3–14, 2015.

Chris Creed and Russell Beale. Abusive interactions with embodied agents. *Interaction Studies*, 9(3):481–503, 2008. ISSN 1572-0373. doi: 10.1075/is.9.3.07cre.

James Cummins. Linguistic interdependence and the educational development of bilingual children. *Review of educational research*, 49(2):222–251, 1979.

Jim Cummins. *Language, power, and pedagogy: Bilingual children in the crossfire*, volume 23. Multilingual Matters, 2000.

Antonella De Angeli and Sheryl Brahnham. I hate you! disinhibition with virtual partners. *Interacting with computers*, 20(3):302–310, 2008.

Antonella De Angeli and Rollo Carpenter. Stupid computer! abuse and social identities. In *Proc. INTERACT 2005 workshop Abuse: The darker side of Human-Computer Interaction*, pages 19–25, 2005.

Antonella De Angeli, Sheryl Brahnham, Peter Wallis, and Alan Dix. Misuse and abuse of interactive technologies. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*, pages 1647–1650. ACM, 2006.

Jill G De Villiers and Valerie E Johnson. The information in third-person/s: acquisition across dialects of american english. *Journal of Child Language*, 34(1):133–158, 2007.

Delpit. *Other People's Children: Cultural Conflict in the Classroom*, volume 46. New Press, 1995. ISBN 1565841794. doi: 10.2307/358724. URL

<http://books.google.com/books?hl=en&lr=id=H9lYESPjKjACpgis=1>.

Kenneth A Dodge, John D Coie, Gregory S Pettit, and Joseph M Price. Peer status and aggression in boys' groups: Developmental and contextual analyses. *Child development*, 61(5):1289–1309, 1990.

Steffi Domagk, Ruth N Schwartz, and Jan L Plass. Interactivity in multimedia learning: An integrated model. *Computers in Human Behavior*, 26(5):1024–1033, 2010.

Sidney DMello and Art Graesser. Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2):145–157, 2012.

Penelope Eckert. *Language variation as social practice: The linguistic construction of identity in Belten High*. Wiley-Blackwell, 2000.

Penelope Eckert. Variation and the indexical field. *Journal of sociolinguistics*, 12(4):453–476, 2008.

Penelope Eckert and John R Rickford. *Style and sociolinguistic variation*. Cambridge University Press, 2001.

Derek Edwards and Neil Mercer. *Common knowledge*, 1987.

Jan Edwards, Megan Gross, Maryellen MacDonald, M Brown, and M Seidenberg. Examining aspects of code-switching ability in children who speak african american english. *Poster presented at SRCLD, Madison, WI*, 2010.

R. Eglash, J. E. Gilbert, V. Taylor, and S. R. Geier. Culturally Responsive Computing in Urban, After-School Contexts: Two Approaches. *Urban Education*, 48(5):629–656, 2013. ISSN 0042-0859. doi: 10.1177/0042085913499211. URL <http://uex.sagepub.com/cgi/doi/10.1177/0042085913499211>.

Christopher Emdin. *For White folks who teach in the Hood... and the rest of y'all too: Reality pedagogy and urban education*. Beacon Press, 2016.

Chrostopher Emdin. "Teaching the Urban Neo-Indiginous": Hip-Hop, Ethnic Divides, and Ur-

- ban Science Education. In *American Education Research Association*, Denver, Colorado, 2010.
- Thomas W Farmer and Philip C Rodkin. Antisocial and prosocial correlates of classroom social positions: The social network centrality perspective. *Social Development*, 5(2):174–188, 1996.
- Daniel C Feldman. The development and enforcement of group norms. *Academy of management review*, 9(1):47–53, 1984.
- Roberta A Ferrara, Ann L Brown, and Joseph C Campione. Children’s learning and transfer of inductive reasoning rules: Studies of proximal development. *Child development*, pages 1087–1099, 1986.
- Joshua A Fishman. *Language and ethnicity in minority sociolinguistic perspective*. Multilingual Matters, 1989.
- H Fogel and Lc Ehri. Teaching Elementary Students Who Speak Black English Vernacular to Write in Standard English: Effects of Dialect Transformation Practice. *Contemporary educational psychology*, 25(2):212–235, 4 2000. ISSN 0361-476X. doi: 10.1006/ceps.1999.1002. URL <http://www.sciencedirect.com/science/article/pii/S0361476X99910020>.
- Signithia Fordham. Speaking standard english from nine to three. *Kids talk: Strategic language use in later childhood*, pages 205–216, 1998.
- Signithia Fordham. Dissin’ the standard: Ebonics as guerrilla warfare at capital high. *Anthropology & Education Quarterly*, 30(3):272–293, 1999.
- Signithia Fordham and John U Ogbu. Black students’ school success: Coping with the burden of acting white. *The urban review*, 18(3):176–206, 1986.
- Carmen Fought. A majority sound change in a minority community:/u/-fronting in chicano english. *Journal of sociolinguistics*, 3(1):5–23, 1999.
- Sandra Fradd and Okhee Lee. Science for all: a promise or a pipe dream for bilingual students?

19(2):261–278, 1995.

Brandi N Frisby and Matthew M Martin. Instructor–student and student–student rapport in the classroom. *Communication Education*, 59(2):146–164, 2010.

Karen Gallas. *Talking their way into science: Hearing children’s questions and theories, responding with curricula*. Teachers College Press, 1995.

James Garbarino. Lost boys: Why our sons turn to violence and how we can save them. *Reaching Today’s Youth: The Community Circle of Caring Journal*, 3(4):7–10, 1999.

Eugene E Garcia and Delis Cuellar. Who are these linguistically and culturally diverse students?. *Teachers College Record*, 108(11):2220–2246, 2006.

Paul J Germann and Roberta J Aram. Student performances on the science processes of recording data, analyzing data, drawing conclusions, and providing evidence. *Journal of research in science teaching*, 33(7):773–798, 1996.

Howard Giles and Mikaela L Marlow. Theorizing language attitudes existing frameworks, an integrative model, and new directions<sup>1</sup>. *Annals of the International Communication Association*, 35(1):161–197, 2011.

Henry A Giroux. *Teachers as intellectuals: Toward a critical pedagogy of learning*. Greenwood Publishing Group, 1988.

Amanda Godley and Allison Escher. Bidialectal African American adolescents’ beliefs about spoken language expectations in English classrooms. *Journal of Adolescent and Adult Literacy*, 55(8):704–713, 5 2012. ISSN 10813004. doi: 10.1002/JAAL.00085. URL <http://doi.wiley.com/10.1002/JAAL.00085>.

Amanda J Godley and Angela Minnici. Critical language pedagogy in an urban high school english class. *Urban Education*, 43(3):319–346, 2008.

Norma González, Rosi Andrade, Marta Civil, and Luis Moll. Bridging funds of distributed knowledge: Creating zones of practices in mathematics. *Journal of Education for students*

*placed at risk*, 6(1-2):115–132, 2001.

Arthur C Graesser, Moongee Jeon, and David Dufty. Agent technologies designed to facilitate interactive knowledge construction. *Discourse processes*, 45(4-5):298–322, 2008.

Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. Creating rapport with virtual agents. In *International Workshop on Intelligent Virtual Agents*, pages 125–138. Springer, 2007.

Judith L Green and Carol N Dixon. Talking knowledge into being: Discursive and social practices in classrooms. *Linguistics and education*, 5(3-4):231–239, 1993.

Lisa J Green. *African American English: a linguistic introduction*. Cambridge University Press, 2002.

Vera F Gutiérrez-Clellen, Janet Calderón, and Susan Ellis Weismer. Verbal working memory in bilingual children. *Journal of Speech, Language, and Hearing Research*, 47(4):863–876, 2004.

Magnus Haake and Agneta Gulz. Visual stereotypes and virtual pedagogical agents. *Journal of Educational Technology & Society*, 11(4):1, 2008.

MAK Halliday and JR Martin. General orientation. *Writing science: Literacy and discursive power*, pages 2–24, 1993.

Beth Harry, JANETTE K Klingner, KEITH M Sturges, and ROBERT F Moore. Of rocks and soft places: Using qualitative methods to investigate disproportionality. *Racial inequity in special education*, 7192, 2002.

Willard W Hartup. The company they keep: Friendships and their developmental significance. *Child development*, 67(1):1–13, 1996.

Shirley Brice Heath. *Ways with words: Language, life and work in communities and classrooms*. cambridge university Press, 1983.

Shirley Brice Heath. Oral and literate traditions among black americans living in poverty. *Amer-*

- ican Psychologist*, 44(2):367, 1989.
- Steffi Heidig and Geraldine Clarebout. Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review*, 6(1):27–54, 2011.
- Lyn Henderson. Instructional Design of Interactive Multimedia: A Cultural Critique. 44(4): 85–104, 2013.
- Dorothy Holland, William Lachicotte, Debra Skinner, and Carol Cain. Agency and identity in cultural worlds, 1998.
- David F Jackson, Billie Jean Edwards, and Carl F Berger. Teaching the design and interpretation of graphs through computer-aided graphical data analysis. *Journal of Research in Science Teaching*, 30(5):483–501, 1993.
- Christopher Jencks and Meredith Phillips. The black-white test scope gap: Why it persists and what can be done. *The Brookings Review*, 16(2):24–27, 1998.
- Amy M Johnson, Matthew D DiDonato, and Martin Reisslein. Animated agents in k-12 engineering outreach: Preferred agent characteristics across age levels. *Computers in Human Behavior*, 29(4):1807–1815, 2013a.
- Amy M Johnson, Gamze Ozogul, Roxana Moreno, and Martin Reisslein. Pedagogical agent signaling of multiple visual engineering representations: The case of the young female agent. *Journal of Engineering Education*, 102(2):319–337, 2013b.
- W Lewis Johnson, Jeff W Rickel, James C Lester, et al. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial intelligence in education*, 11(1):47–78, 2000.
- Martin H Jones, Shannon R Audley-Piotrowski, and Sarah M Kiefer. Relationships among adolescents’ perceptions of friends’ behaviors, academic self-concept, and math performance. *Journal of Educational Psychology*, 104(1):19, 2012.
- Sara Kiesler and Lee Sproull. social human-computer interaction. In *Human values and the*

- design of computer technology*, pages 191–199. Center for the Study of Language and Information, 1997.
- Yanghee Kim and Amy L Baylor. A social-cognitive framework for pedagogical agents as learning companions. *Educational Technology Research and Development*, 54(6):569–596, 2006.
- Yanghee Kim and Quan Wei. The impact of learner attributes and learner choice in an agent-based environment. *Computers & Education*, 56(2):505–514, 2011.
- Yanghee Kim, Amy L Baylor, and E Shen. Pedagogical agents as learning companions: the impact of agent emotion and gender. *Journal of Computer Assisted Learning*, 23(3):220–234, 2007.
- Alison King. Verbal interaction and problem-solving within computer-assisted cooperative learning groups. *Journal of Educational Computing Research*, 5(1):1–15, 1989.
- Alison King. Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal*, 27(4):664–687, 1990.
- Alison King. Effects of training in strategic questioning on children’s problem-solving performance. *Journal of Educational Psychology*, 83(3):307, 1991.
- Alison King. Facilitating Elaborative Learning Through Guided Student-Generated Questioning. *Educational Psychologist*, 27(1):111–126, 1992.
- Alison King. Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American educational research journal*, 31(2):338–368, 1994.
- Alison King. Ask to think-tel why: A model of transactive peer tutoring for scaffolding higher level complex learning. *Educational psychologist*, 32(4):221–235, 1997.
- Alison King. Chapter 2 SCRIPTING COLLABORATIVE LEARNING PROCESSES : A COGNITIVE PERSPECTIVE. 2002.
- Lisa M Koch, Alan M Gross, and Russell Kolts. Attitudes toward black english and code switch-

ing. *Journal of Black psychology*, 27(1):29–42, 2001.

Nicole C Krämer. Social effects of virtual assistants. a review of empirical results with regard to communication. In *International Workshop on Intelligent Virtual Agents*, pages 507–508. Springer, 2008.

Brigitte Krenn, Birgit Endrass, Felix Kistler, and Elisabeth André. Effects of language variety on personality perception in embodied conversational agents. In *International Conference on Human-Computer Interaction*, pages 429–439. Springer, 2014.

Anthony Kroch and William Labov. Linguistic society of america: resolution in response to arthur jensen (1969). *Linguistic Society of American Bulletin (March)*. Washington: Linguistic Society of America, 1972.

Vivien Kuhne, Astrid Marieke Rosenthal-Von Der Putten, and Nicole C. Kramer. Using linguistic alignment to enhance learning experience with pedagogical agents: The special case of dialect. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8108 LNAI:149–158, 2013. ISSN 03029743. doi: 10.1007/978-3-642-40415-3\_13.

G. Ladson-Billings. Toward a Theory of Culturally Relevant Pedagogy. *American Educational Research Journal*, 32(3):465–491, 1995a. ISSN 0002-8312. doi: 10.3102/00028312032003465.

Gloria Ladson-Billings. Toward a theory of culturally relevant pedagogy. *American educational research journal*, 32(3):465–491, 1995b.

Gloria Ladson-Billings. Just what is critical race theory and what’s it doing in a nice field like education? *International Journal of Qualitative Studies in Education*, 11(1):7–24, 1998. ISSN 0951-8398. doi: 10.1080/095183998236863.

Magdalene Lampert. When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American educational research journal*, 27(1):29–63,



1990.

Sonja L Lanehart. African american vernacular english and education: The dynamics of pedagogy, ideology, and identity. *Journal of English linguistics*, 26(2):122–136, 1998.

Bibb Latane. The psychology of social impact. *American psychologist*, 36(4):343, 1981.

Jean Lave and Etienne Wenger. Communities of practice. *Retrieved June*, 9:2008, 1998.

O. Lee, H. Quinn, and G. Valdes. Science and Language for English Language Learners in Relation to Next Generation Science Standards and With Implications for Common Core State Standards for English Language Arts and Mathematics. *Educational Researcher*, 42(4):222–233, 2013.

Okhee Lee. Equity for linguistically and culturally diverse students in science education: A research agenda. *Teachers College Record*, 105(3):465–489, 2003.

Okhee Lee and Sandra H Fradd. Science for All, including Students from Non-English-Language Backgrounds. *Educational Researcher*, 27(4):12–21, 1998. ISSN 0013189X, 1935102X. doi: 10.3102/0013189X027004012. URL <http://www.jstor.org/stable/1176619>.

Okhee Lee, Rachael A Deaktor, Juliet E Hart, Peggy Cuevas, and Craig Enders. An instructional intervention’s impact on the science and literacy achievement of culturally and linguistically diverse elementary students. *Journal of Research in Science Teaching*, 42(8):857–887, 2005.

Jay L Lemke. *Talking science: Language, learning, and values*. ERIC, 1990.

Steven D Levitt and Roland G Fryer. Falling behind: New evidence on the black-white achievement gap. *Education Next*, 4(4):64–71, 2004.

Max M Louwerse, Arthur C Graesser, Shulan Lu, and Heather H Mitchell. Social cues in animated conversational agents. *Applied Cognitive Psychology*, 19(6):693–704, 2005.

Michael Madaio, Justine Cassell, and Amy Ogan. The impact of peer tutors use of indirect feedback and instructions. Philadelphia, PA: International Society of the Learning Sciences., 2017.

- Michael A Madaio, Amy Ogan, and Justine Cassell. The effect of friendship and tutoring roles on reciprocal peer tutoring strategies. In *International Conference on Intelligent Tutoring Systems*, pages 423–429. Springer, 2016.
- Christine Mallinson and Becky Child. Communities of practice in sociolinguistic description: Analyzing language and identity practices among black women in appalachia. *Gender & Language*, 1(2), 2007.
- Mark Mason. Critical thinking and learning. *Educational philosophy and theory*, 39(4):339–349, 2007.
- Richard E Mayer. 14 principles based on social cues in multimedia learning: Personalization voice image and embodiment principles. *The Cambridge handbook of multimedia learning*, page 345, 2014a.
- Richard E Mayer. *The Cambridge handbook of multimedia learning, 2nd edition*. Cambridge university press, 2014b.
- Richard E Mayer, Kristina Sobko, and Patricia D Mautone. Social cues in multimedia learning: Role of speaker’s voice. *Journal of Educational Psychology*, 95(2):419, 2003.
- Terry Meier. The case for ebonics as part of exemplary teacher preparation. *Making the connection: Language and academic achievement among African American students*, pages 97–114, 1999.
- Neil Mercer and Lyn Dawes. The value of exploratory talk. *Exploring talk in school*, pages 55–71, 2008.
- Neil Mercer, Lyn Dawes, and Judith Kleine Staarman. Dialogic teaching in the primary science classroom. *Language and Education*, 23(4):353–369, 2009.
- Miriam Meyerhoff. Dealing with gender identity as a sociolinguistic variable. *Rethinking language and gender research: Theory and practice*, pages 202–27, 1996.
- Sarah Michaels. ”Sharing time”: Children’s narrative styles and differential access

- to literacy. *Language in Society*, 10(3):423–442, 1981. ISSN 0047-4045. doi: 10.1017/S0047404500008861. URL <http://www.jstor.org/stable/4167263>.
- Sarah Michaels. Listening and Responding: Hearing in Children’s Classroom Narratives. *Theory into Practice*, 23(3):218–224, 2013.
- G. Mohatt and F. Erickson. Cultural Differences in Teaching Styles in an Odawa School: A Sociolinguistic Approach. *Culture and the Bilingual Classroom: Studies in Classroom Ethnography*, 105, 1981.
- Roxana Moreno and Terri Flowerday. Students’ choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. *Contemporary educational psychology*, 31(2):186–207, 2006. doi: 10.1016/j.cedpsych.2005.05.002.
- Roxana Moreno, Richard E Mayer, Hiller A Spires, and James C Lester. The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and instruction*, 19(2):177–213, 2001.
- Keith E Nainby, John T Warren, and Christopher Bollinger. Articulating contact in the classroom: Towards a constitutive focus in critical pedagogy. *Language and Intercultural Communication*, 3(3):198–212, 2003.
- Clifford Nass, Byron Reeves, and Glenn Leshner. Technology and roles: A tale of two tvs. *Journal of Communication*, 46(2):121–128, 1996.
- Clifford Nass, Katherine Isbister, and Eun-ju Lee. Truth is Beauty: Researching Embodied Conversational Agents. *Embodied Conversational Agents*, pages 374–402, 2000.
- National Center for Education Statistics. The condition of education 2015. <https://nces.ed.gov/pubs2015/2015144.pdf>, May 2015.
- National Center for Education Statistics. Status and trends in the education of racial and ethnic groups 2017. <https://nces.ed.gov/pubs2017/2017051.pdf>, July 2017.
- National Research Council. *A Framework for K-12 Science Education: Practices, Crosscutting*

- Concepts, and Core Ideas*, volume 23. The National Academies Press, Washington, DC, 2012.
- Richard S Newman. How self-regulated learners cope with academic difficulty: The role of adaptive help seeking. *Theory into practice*, 41(2):132–138, 2002.
- Sonia Nieto. What does it mean to affirm diversity?. *School Administrator*, 56(6):32–35, 1999.
- Sonia Nieto. *The light in their eyes: Creating multicultural learning communities*. Teachers College Press, 2015.
- Stephen P Norris and Linda M Phillips. How literacy in its fundamental sense is central to scientific literacy. *Science education*, 87(2):224–240, 2003.
- Elinor Ochs. Constructing social identities: a language socialization perspective. *Research on Language and Social Interaction*, 26(3):287–306, 1993.
- Elinor Ochs. Constructing social identity: A language socialization perspective. *Intercultural discourse and communication: The essential readings*, pages 78–91, 2005.
- J. U. Ogbu. Beyond Language: Ebonics, Proper English, and Identity in a Black-American Speech Community. *American Educational Research Journal*, 36(2):147–184, 1999. ISSN 0002-8312. doi: 10.3102/00028312036002147.
- John U Ogbu. Minority coping responses and school experience. *The Journal of Psychohistory*, 18(4):433, 1991.
- Gamze Ozogul, Amy M Johnson, Robert K Atkinson, and Martin Reisslein. Investigating the impact of pedagogical agent gender matching and learner choice on learning outcomes and perceptions. *Computers & Education*, 67:36–50, 2013.
- Aannemarie Sullivan Palinscar and Ann L Brown. Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and instruction*, 1(2):117–175, 1984.
- Judy M Parr and Michael AR Townsend. Environments, processes, and mechanisms in peer learning. *International journal of educational research*, 37(5):403–423, 2002.

- Sabine Payr. The virtual university's faculty: An overview of educational agents. *Applied Artificial Intelligence*, 17(1):1–19, 2003.
- Reinhard Pekrun and Lisa Linnenbrink-Garcia. Academic emotions and student engagement. In *Handbook of research on student engagement*, pages 259–282. Springer, 2012.
- Anthony D Pellegrini. A longitudinal study of boys' rough-and-tumble play and dominance during early adolescence. *Journal of Applied Developmental Psychology*, 16(1):77–93, 1995.
- Jennifer Renn. Acquiring style: the development of dialect shifting among african american children. *Analysis*, pages 1–129, 2010. ISSN 04194209.
- Lauren Resnick, Christa Asterhan, and Sherice Clarke. *Socializing intelligence through academic talk and dialogue*. American Educational Research Association, 2015.
- John M Reveles, Ralph Cordova, and Gregory J Kelly. Science literacy and academic identity formulation. *Journal of Research in Science Teaching*, 41(10):1111–1144, 2004.
- John Russell Rickford. *African American vernacular English: Features, evolution, educational implications*. Wiley-Blackwell, 1999.
- Jeanene Robinson and Mia Biran. Discovering self: Relationships between african identity and academic achievement. *Journal of Black Studies*, 37(1):46–68, 2006.
- Barbara Rogoff. Apprenticeships in thinking. *Cognitive development in social context*. Oxford, England: Oxford University Press, 1990.
- Carolyn Rosé, Yi-Chia Wang, Yue Cui, Jaime Arguello, Karsten Stegmann, Armin Weinberger, and Frank Fischer. Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International journal of computer-supported collaborative learning*, 3(3):237–271, 2008.
- Milton E Rosenbaum. The repulsion hypothesis: On the nondevelopment of relationships. *Journal of Personality and Social Psychology*, 51(6):1156, 1986.
- Rinat B Rosenberg-Kima, Amy L Baylor, E Ashby Plant, and Celeste E Doerr. Interface agents

- as social models for female students: The effects of agent visual presence and appearance on female students attitudes and beliefs. *Computers in Human Behavior*, 24(6):2741–2756, 2008.
- Richard Rothstein. *Class and schools*. Teachers College, Columbia University, 2004.
- Russell W Rumberger and Gregory J Palardy. Does segregation still matter? the impact of student composition on academic achievement in high school. *Teachers college record*, 107(9):1999, 2005.
- Nikol Rummel, Hans Spada, and Sabine Hauser. Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning*, 4(1):69–92, 2009.
- Noah L Schroeder and Olusola O Adesope. A systematic review of pedagogical agents persona, motivation, and cognitive load implications for learners. *Journal of Research on Technology in Education*, 46(3):229–251, 2014.
- Noah L Schroeder and Chad M Gotch. Persisting issues in pedagogical agent research. *Journal of Educational Computing Research*, 53(2):183–204, 2015.
- Noah L Schroeder, Olusola O Adesope, and Rachel Barouch Gilbert. How effective are pedagogical agents for learning? a meta-analytic review. *Journal of Educational Computing Research*, 49(1):1–39, 2013.
- Noah L Schroeder, William L Romine, and Scotty D Craig. Measuring pedagogical agent persona and the influence of agent persona on learning. *Computers & Education*, 109:176–186, 2017.
- Philip H Scott, Eduardo F Mortimer, and Orlando G Aguiar. The tension between authoritative and dialogic discourse: A fundamental characteristic of meaning making interactions in high school science lessons. *Science Education*, 90(4):605–631, 2006.
- Gale Seiler. Reversing the standard direction: Science emerging from the lives of african american students. *Journal of Research in Science teaching*, 38(9):1000–1014, 2001.

- J Siegel. Pidgins, creoles, and minority dialect in education. *Concise Encyclopedia of Sociolinguistics*, Elsevier, Oxford, pages 747–749, 2001.
- Jeff Siegel. *Applied Creolistics in the 21st century*. na, 2002.
- Jeff Siegel. Keeping Creoles and Dialects out of the Classroom: is it justified? In Shondel Nero, editor, *Dialects, Englishes, Creoles, and Education*, pages 39–67. Psychology Press, 2006a. ISBN 0805846581. doi: 10.4324/9780203928660. URL <http://books.google.com/books?hl=en&r=id=SBIutJgpNacCpgis=1>.
- Jeff Siegel. Language ideologies and the education of speakers of marginalized language varieties: Adopting a critical awareness approach. *Linguistics and Education*, 17(2):157–174, 2006b.
- Jeff Siegel. *The emergence of pidgin and creole languages*. Oxford University Press, 2008.
- Jeff Siegel. *Second dialect acquisition*. Cambridge University Press, 2010.
- Ramadhar Singh and Soo Yan Ho. Attitudes and attraction: A new test of the attraction, repulsion and similarity-dissimilarity asymmetry hypotheses. *British Journal of Social Psychology*, 39(2):197–211, 2000.
- Elizabeth Sklar and Debbie Richards. Agent-based systems for human learners. *The Knowledge Engineering Review*, 25(2):111–135, 2010.
- Geneva Smitherman. *Talkin and testifyin: The language of Black America*, volume 51. Wayne State University Press, 1977.
- Daniel Solorzano, Miguel Ceja, and Tara Yosso. Critical race theory, racial microaggressions, and campus racial climate: The experiences of african american college students. *Journal of Negro Education*, pages 60–73, 2000.
- RA Sprott. The development of conversational skills: contextual and functional analyses of children’s discourse markers. In *SRCD Biennial Meeting, New Orleans*, 1993.
- Lee Sproull, Mani Subramani, Sara Kiesler, Janet H Walker, and Keith Waters. When the inter-

- face is a face. *Human-Computer Interaction*, 11(2):97–124, 1996.
- Claude Steele. *Stereotype threat and African-American student achievement*. na, 2003.
- Claude M Steele. A threat in the air: How stereotypes shape intellectual identity and performance. *American psychologist*, 52(6):613, 1997.
- Jeremiah Sullins, Scotty D Craig, and Arthur C Graesser. Tough love: The influence of an agent’s negative affect on students’ learning. In *AIED*, pages 677–679, 2009.
- Julie Sweetland. *Teaching writing in the African American classroom: a sociolinguistic approach*. PhD thesis, Stanford University, 2006.
- Julie Sweetland. Making dissatisfaction divine: An inspired approach to whole-school reform. *International Journal of Whole Schooling*, 4(1):13–21, 2008.
- John Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2):257–285, 1988.
- Susan R Swing and Penelope L Peterson. The relationship of student ability and small-group interaction to student achievement. *American Educational Research Journal*, 19(2):259–274, 1982.
- Hanni U Taylor. *Standard English, black English, and bidialectalism: a controversy*. Peter Lang Pub Inc, 1989.
- N. P. Terry. Examining Relationships Among Dialect Variation and Emergent Literacy Skills. *Communication Disorders Quarterly*, 33(2):67–77, 2010. ISSN 1525-7401. doi: 10.1177/1525740110368846. URL <http://cdq.sagepub.com/cgi/doi/10.1177/1525740110368846>.
- Nicole Patton Terry and Hollis S Scarborough. The phonological hypothesis as a valuable framework for studying the relation of dialect variation to early reading skills. 2011.
- Nicole Patton Terry, Carol McDonald Connor, Shurita Thomas-Tate, and Michael Love. Examining relationships among dialect variation, literacy skills, and school context in first grade.



- Journal of Speech, Language, and Hearing Research*, 53(1):126–145, 2010.
- By States The Next Generation Science Standards For States. Next Generation Science Standards (“NGSS”), 2014. URL <http://www.nextgenscience.org>.
- Linda Tickle-Degnen and Robert Rosenthal. The nature of rapport and its nonverbal correlates. *Psychological inquiry*, 1(4):285–293, 1990.
- Jessica Twyford and Scotty D Craig. Modeling goal setting within a multimedia environment on complex physics content. *Journal of Educational Computing Research*, 55(3):374–394, 2017.
- United States Census Bureau. Most children younger than age 1 are minorities. <https://www.census.gov/newsroom/releases/archives/population/cb12-90.html>, May 2012.
- United States Census Bureau. Income and poverty in the united states: 2016. <https://census.gov/content/dam/Census/library/publications/2017/demo/P60-259.pdf>, September 2017.
- United States Congress. National education goals. <http://uscode.house.gov/statviewer.htm?volume=108page=130> March 1994.
- Hans van der Meij, Jan van der Meij, and Ruth Harmsen. Animated pedagogical agents effects on enhancing student motivation and learning in a science inquiry learning environment. *Educational technology research and development*, 63(3):381–403, 2015.
- Marieke van Heugten, Dena R Krieger, and Elizabeth K Johnson. The developmental trajectory of toddlers comprehension of unfamiliar regional accents. *Language Learning and Development*, 11(1):41–65, 2015.
- J. Van Hofwegen and W. Wolfram. Coming of Age in African American English: A Longitudinal Study. *Journal of Sociolinguistics*, 14(4):427–455, 2010.
- Janneke Van Hofwegen. Dyadic analysis: Factors affecting african american english usage and accommodation in adolescent peer dyads. *Language & Communication*, 41:28–45, 2015.
- K VanLehn, A Graesser, GT Jackson, P Jordan, A Olney, and CP Rosé. Natural language tutor-

- ing: A comparison of human tutors, computer tutors, and text. 2007.
- Maria Varelas, Joe Becker, Barbara Luster, and Stacy Wenzel. When genres meet: Inquiry into a sixth-grade urban science class. *Journal of Research in Science Teaching*, 39(7):579–605, 2002.
- Paul Vedder. *Cooperative learning: A study on processes and effects of cooperation between primary school children*, volume 38. Taylor & Francis Group, 1985.
- George Veletsianos. Cognitive and affective benefits of an animated pedagogical agent: Considering contextual relevance and aesthetics. *Journal of Educational Computing Research*, 36(4): 373–377, 2007.
- George Veletsianos. Contextually relevant pedagogical agents: Visual appearance, stereotypes, and first impressions and their impact on learning. *Computers & Education*, 55(2):576–585, 2010.
- George Veletsianos, Cassandra Scharber, and Aaron Doering. When sex, drugs, and violence enter the classroom: Conversations between adolescents and a female pedagogical agent. *Interacting with Computers*, 20(3):292–301, 2008. ISSN 09535438. doi: 10.1016/j.intcom.2008.02.007.
- Lev S Vygotsky. *Sobranie sochinenii. Moscú: Pedagogika*, 1982.
- Peter Wallis. Believable conversational agents: Introducing the intention map. In *International conference on autonomous agents and multiagent systems*, pages 17–22, 2005.
- William Yang Wang, Samantha Finkelstein, Amy Ogan, Alan W Black, and Justine Cassell. Love ya, jerkface: using sparse log-linear models to build positive (and impolite) relationships with teens. In *Proceedings of the 13th annual meeting of the special interest group on discourse and dialogue*, pages 20–29. Association for Computational Linguistics, 2012.
- Beth Warren, Cynthia Ballenger, Mark Ogonowski, Ann S Rosebery, and Josiane Hudicourt-Barnes. Rethinking diversity in learning science: The logic of everyday sense-making. *Journal*

- of research in science teaching*, 38(5):529–552, 2001.
- Lyn Webb and Paul Webb. Introducing Discussion into Multilingual Mathematics Classrooms : An Issue of Code Switching ? 32(June):26–32, 2008.
- Noreen M Webb. Peer interaction and learning in small groups. *International journal of Educational research*, 13(1):21–39, 1989.
- Gordon Wells. *Dialogic inquiry: Towards a socio-cultural practice and theory of education*. Cambridge University Press, 1999.
- Etienne Wenger. *Communities of Practice: Learning, Meaning, and Identity*, volume 9. 1998. ISBN 0521663636. doi: 10.2277/0521663636.
- James V Wertsch, Peter Tulviste, and Fran Hagstrom. A sociocultural approach to agency. *Contexts for learning: Sociocultural dynamics in childrens development*, 23:336–356, 1993.
- R. S. Wheeler and Rebecca Swords. Codeswitching: Tools of Language and Culture Transform the Dialectally Diverse Classroom. *Language Arts*, 81(6):470–480, 2004.
- Walt Wolfram and Donna Christian. *Dialects and education: Issues and answers*. Prentice Hall, 1989.
- Huay Lit Woo. Designing multimedia learning environments using animated pedagogical agents: factors and issues. *Journal of Computer Assisted Learning*, 25(3):203–218, 2009.
- Jack C Wright, Mary Giammarino, and Harry W Parad. Social status in small groups: Individual–group similarity and the social” misfit.”. *Journal of Personality and Social Psychology*, 50(3): 523, 1986.
- Erna Yackel, Paul Cobb, Terry Wood, Grayson Wheatley, and Graceann Merkel. The importance of social interaction in childrens construction of mathematical knowledge. *Teaching and learning mathematics in the*, 19903, 1990.
- Sean Zdenek. A new agenda in (critical) discourse analysis: Theory, methodology and interdisciplinarity. *Language in Society*, 35(4):613–617, 2006.