Boston College
Lynch School of Education and Human Development

Department of Measurement, Evaluation, Statistics, and Assessment

MEASURING STUDENTS' PERCEPTIONS OF FACULTY AVAILABILITY OUTSIDE OF CLASS USING RASCH/GUTTMAN SCENARIO SCALES

Dissertation

by

KATHERINE ANN REYNOLDS

submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

May 2020

ABSTRACT

MEASURING STUDENTS' PERCEPTIONS OF FACULTY AVAILABILITY OUTSIDE OF CLASS USING RASCH/GUTTMAN SCENARIO SCALES

Katherine A. Reynolds, Author

Larry H. Ludlow, Chair

Interaction with faculty is one of the most important aspects of completing an undergraduate degree (Chambliss & Takacs, 2014). At traditional colleges and universities, much of this interaction takes place within the classroom. However, out-of-class communication (OCC) is also an important part of the college-going experience. Participation in OCC has been associated with many positive undergraduate outcomes, such as motivation (Komarraju et al., 2010) and course grades (Micari & Pazos, 2012).

Prior measurement instruments related to OCC suffered from limitations with respect to construct definition and methodological procedures that limit the interpretability and utility of the scores they yield. My dissertation ameliorates these issues in constructing a new instrument that measures students' perceptions of faculty availability outside of class. This instrument is built using Rasch/Guttman Scenario (RGS) scale methodology, which brings together the frameworks of Rasch measurement (Rasch, 1960/80) and Guttman facet theory design (Guttman, 1954; Guttman 1959). Two scales, each containing seven short scenarios that function as items, were constructed: the Physical Accessibility Scale (PAS) and the Social Engagement Scale (SES). Together, these two scales comprise the Out-of-Class Availability Scales (OCAS). Three facets of physical accessibility and social engagement are identified and represented within the items: arranged meetings, chance encounters, and email.

The OCAS development process and analysis results presented within my dissertation suggest that the RGS methodology is useful for capturing students' perceptions of faculty

availability outside of class. The OCAS can also be used by others to conduct future research on the topic of OCC. Because they measure students' perceptions of availability and not frequency of OCC, the OCAS have value as a potential faculty evaluation tool. Even if students choose not to interact with a particular faculty member outside of class, they would still ideally find that faculty member available for such interaction should the need arise. Finally, the RGS scale development process ensures that OCAS scores are accompanied by qualitative descriptions, which enhances their utility and measurement value.

ACKNOWLEDGMENTS

Although my name is the one that appears on this dissertation, I could not have completed this work without the support, and encouragement of others along the way. I am so grateful for everyone that has helped me along this doctoral journey and fully acknowledge that this page is not enough to express my gratitude.

Thank you to my dissertation chair, Dr. Larry Ludlow. You are a wonderful example of a faculty member who is both accessible for and engaged in interactions with students outside of class. I also want to thank my other dissertation committee members, Dr. Jessica Greene, and Dr. Ana Martinez-Alemán. Your insightful feedback and support has been instrumental throughout this process.

Thank you to the Boston College Office of Institutional Research and Planning for assisting me with participant recruitment and seeing the value in this study. I also appreciate the assistance of my good friend Dr. David Brown in obtaining a sample for my pre-pilot.

I will never be able to express sufficient gratitude to my parents, John and Susan Reynolds, and my sister Dana. I appreciate the hours you have spent listening to me talk about this project and for your unwavering support as I pursued my graduate degree.

I have also been lucky to have the support of many wonderful friends. Thank you to Jenny Liu and Caroline Prendergast, who are the best cheerleaders and support system that anyone could ask for. Thank you to Sebastian Moncaleano for always lending an ear throughout this process and pushing me to be a better researcher and analyst. Thank you to all of the other friends I have made within the MESA department; I know all of you will go on to do great things. I am also grateful for my closest friends outside of Boston College who have supported me from afar: Morgan Wright, Danielle Smith, and Kylie Whitlock. Thank you all for never letting me forget that there’s a world beyond graduate school.

Finally, I want to thank my husband, Joseph Sherman. I could not have made it through this PhD without you. Everything I do is built on our foundation and I can’t wait to see where we go next.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

The Problem

Interpersonal relationships are at the heart of the college-going experience for today's undergraduate students and some of the most important relationships students will develop are those with faculty at their institutions (Chambliss & Takacs, 2014). These relationships are complex and multi-faceted, as faculty can play many roles for the students they teach, including adviser, counselor, teacher, or friend (Wilson et al., 1974). For traditional students enrolled at residential undergraduate institutions, these relationships typically begin within the context of structured class time. This is where students meet faculty members and is also likely to be where students will spend the most time interacting with them. However, interacting outside the confines of structured course space is what allows the student-faculty relationship to evolve beyond formal titles, allowing faculty members to fill the roles listed above.

Interaction between college students and faculty outside of class time (termed out-of-class communication, or OCC) has been of interest to researchers for decades (e.g., Pascarella, 1980; Pascarella et al., 1978; Snow, 1973; Theophilides & Terenzini, 1981). OCC has been studied from a variety of perspectives, including examination of the relationship between OCC and student outcomes, as well as how well different student or faculty characteristics can predict the occurrence of OCC. Despite ongoing interest in the topic, the scholarly community has not settled upon a single definition of what constitutes OCC. There is no definitive reason for this lack of a common definition; however, Goldman et al. (2016) suggest it may be because OCC is of interest in multiple fields (e.g., education, communication), each of which have their own standard research practices and definitions. The following list provides a sample of OCC definitions from existing literature:

- "…structured and unstructured activities or conditions that are not directly part of an institution's formal, course-related, instructional processes" (Terenzini et al., 1996, p. 150)
- "…formal and/or informal interaction between faculty and students which takes place outside of formal classrooms and during times other than when class is scheduled" (Zhang, 2006, p. 34)
- "Informal faculty-student interaction that occurs beyond the realm of formal in-class instruction… a wide variety of informal faculty-student contact such as that which occurs before and after class, in or outside of the physical classroom setting, spontaneously on campus, during official office hours, by appointment, or via technological mediums such as the telephone or internet" (Bippus et al., 2001, p. 16)

These three sample definitions are a good representation of the different ways in which researchers across fields have conceptualized OCC. Some, such as in the first definition, are very broad, and provide little clarification regarding OCC that is not already apparent from the phrase itself. The second definition attempts to give some specificity by delineating types of interaction (formal vs. informal) and a designation of time and space (outside of class and not when class is meeting). Lastly, the third definition is a good representation of the level of specificity employed in many individual studies, which attend not only to the time and place of OCC, but also to the various modes through which OCC might occur. While this definition is representative in terms of the level of specificity, it is not representative of any single commonly-accepted OCC definition.

The myriad ways in which OCC has been defined make it challenging to determine the exact nature of relationships between OCC and other variables of interest. Mayhew et al. (2016) allude to this point, noting that “research has varied considerably in how interactions with faculty have been operationalized” (p. 77). The authors suggest this as one plausible explanation for the conflicting findings regarding the effects of OCC for students (reviewed in Chapter 2); some studies report statistically significant positive associations between frequency of OCC and various student outcomes, while others do not consistently demonstrate meaningful relationships, especially when accounting for other components of the undergraduate experience. Faranda (2015) echoes the sentiments of Mayhew et al. (2016), saying that “no uniform agreement exists on the scope of OCC or how best to measure it” (p. 85).

Inconsistencies in OCC definition lead to inconsistences in OCC measurement. Goldman et al. (2016) wrestle with inconsistent measures in their meta-analysis of studies linking OCC to student outcomes. Focusing exclusively on face-to-face OCC, they examined 14 studies that met the following criteria:

- Inclusion of a quantitative measure of OCC,
- Inclusion of a quantitative measure of student affective or cognitive learning, and
- Sufficient statistical detail for the calculation of effect sizes.

The meta-analysis ultimately revealed OCC to have moderate effects on various aspects of students’ affective or cognitive learning. As important to this dissertation are the limitations the authors discuss with respect to their work. They note that fourteen is a small sample size for this kind of meta-analytic research; Goldman et al. (2016) explain that this limitation is due to a dearth of studies meeting their criteria for inclusion. They posit the following:

> Arguably, the biggest reason for this limitation is the variety of incompatible measures that exist to assess OCC throughout the education and communication literature. These instruments vary in their designated focus, making it difficult to interpret the relationship between specific types of OCC and student learning (p. 486).

Despite this conceptual and instrumental diversity, examination of existing OCC-related instruments (discussed at length in Chapter 2) reveals some similarities. First, these measures often address OCC as an observable action that does or does not occur. Instrument items typically ask students to indicate agreement with statements regarding whether or not a specific OCC interaction has or is likely to occur between themselves and a faculty member. These instruments also tend to be similar with respect to their development processes. Those that are designed as scales were all developed using a classical test theory (CTT) approach, relying primarily on factor analytic methods. That said, the specific ways in which these methods are used is not necessarily consistent across instruments—for example, some authors used principal components analysis to develop their measures, while others used factor analysis. Only one study (Cokley et al., 2004) makes a clear distinction between these analytic procedures and presents a rationale for the method selected.

These content- and methodological similarities among existing instruments are suggestive of some general limitations with respect to researchers' current understanding of OCC. Treating OCC as an observable action and focusing on frequency and likelihood of such action is useful for examining the relationships between OCC and other variables (such as was done in the studies reviewed by Goldman et al., (2016)); however such instruments are not capable of demonstrating why OCC occurs (or not). This lack of nuance may explain the contradictory findings of the importance of OCC in current literature. Some qualitative research has sought to explore why students do or do not initiate OCC with faculty members; however, there is currently no quantitative instrument that can shed light on these issues. The pervasiveness of the CTT paradigm among OCC-related instruments also suggests limitations. For reasons that will be discussed in detail in subsequent chapters, scores obtained from these instruments are not accompanied by clear, meaningful interpretations. This renders those scores less useful for informing any sort of action to encourage OCC.

An example illustrates these limitations. Imagine a department chair who is interested in improving OCC between department faculty and students. Gathering data to support this effort, the chair administers a frequency-based, factor-analytic instrument to students enrolled in classes taught by each faculty member. The department chair examines the results of her data collection effort. The scores she has obtained from the students reflect only how often each student interacted with a faculty member outside of class; they reveal nothing about why students felt comfortable initiating these interactions (or not). The department chair cannot qualitatively describe how students see each of her faculty members; nor can she discern clear feedback to provide to her faculty. The department chair is at a loss for how to make meaningful use of the information she has collected and is unclear how to use her data to inform an actionable plan for her department.

Study Purpose and Research Questions

The instrument I develop in this dissertation is a proposed solution to the problem described in the previous paragraph. Rather than measuring on OCC as an action, I focus on a latent (i.e., not directly observable) construct that is salient for OCC across different modes of communication: students' perceptions that faculty are available outside of class. I argue that these perceptions impact students' choices to engage in OCC with particular faculty members, and that measuring them can therefore shed light on students' interactive choices. In addition to this conceptual reframing, I also address the methodological limitations of existing instruments through employment of the Rasch measurement paradigm (described in Chapter 2), specifically employing the Rasch/Guttman Scenario (RGS) scale development approach. Here I discuss my rationale for these choices and present my formal research questions.

Measuring students' perceptions that faculty are available outside of class is beneficial for several reasons. First, measuring this latent construct has the potential to reveal why OCC is relatively uncommon (see Chapter 2 for a discussion of this phenomenon). I reason that students who perceive faculty to be available outside of class will engage in OCC more frequently and that a lack of perceived faculty availability may contribute to relatively low levels of OCC among undergraduate students. Given the fair amount of research supporting that engaging in OCC is typically a positive experience for students (again, see Chapter 2), the higher education community has a vested interest in increasing its occurrence. Measuring students' perceptions of faculty availability may be informative in these efforts. I examine the relationship between students' perceptions of availability and actual participation in OCC as a secondary research question in my dissertation to begin exploration of this utility.

Second, measuring students' perceptions of faculty availability outside of class may be useful for departmental improvement and faculty evaluation. It is unrealistic to expect all students to develop close out-of-class relationships with all faculty they encounter. Ideally, students would still perceive faculty with whom they are not close to be available outside of class should the need for OCC arise. Thus, measuring students' perceptions of availability (a latent construct), in contrast to an observable occurrence such as the number of times students actually met with a faculty member, may be a more valuable indicator of faculty quality. Evaluating this utility is beyond the scope of my dissertation; however, this possibility will be discussed further in the Future Research section of Chapter 5.

I develop my instrument within the Rasch measurement paradigm in order to address the interpretability limitations of existing measures. Specifically, I combine the underlying principles of Rasch measurement (Rasch, 1960/80) with Guttman facet theory design (Guttman, 1954; Guttman, 1959) to create an instrument comprised of short scenarios. Scales constructed in this framework are known as Rasch/Guttman Scenario (RGS) scales (Ludlow et al., in press). This development approach (described in detail in Chapter 3) uses purposeful design throughout all stages to ultimately yield scale scores that have clear interpretations and can provide actionable feedback. Both the methodological approach and the selection of students' perceptions of faculty availability as the construct of interest ameliorate the problems of the hypothetical department chair in the previous section. Were she to administer my dissertation instrument rather than an existing classical test theory-based measure focused on frequency, her end result would be a set of interpretable scores with qualitative descriptions of how students perceive department faculty. This kind of information would enable her to create appropriate and actionable plans to improve student-faculty interaction within the department.

My dissertation's primary research question follows from this discussion:

1. Can Rasch/Guttman Scenario scales provide valid and reliable measurement of student perceptions of faculty availability outside of class?

In addition to this primary research question, I also explore the following secondary research question:

2. What is the relationship between scores from these scales and students' participation in out-of-class communication?

I address my secondary research question using the instrument developed in my first. I examine relationships between scores from my instrument and other OCC-related variables including frequency and satisfaction.

Study Significance

My dissertation makes both substantive and methodological contributions to existing literature. Substantively, my focus on students' perceptions of faculty availability adds to ongoing research conversations regarding OCC and its importance to undergraduate student outcomes. Once fully validated, my instrument will allow future researchers to continue exploration of OCC in a more nuanced way that attends to students' perceptions of faculty availability. The finalized instrument may also be of use to those working on faculty development in higher education, such as department heads or those employed in teaching centers.

My instrument has strong evidence for construct validity because of the rigorous development approach employed. RGS scales require a thorough and iterative process of construct definition, with many refinements made throughout the procedures. There is reason to be confident that my instrument adequately attends to the complexity of students' perceptions of faculty availability outside of class and captures that complexity in a faithful way. This enhances the utility of my instrument for the future researchers to whom I allude in the previous paragraph.

Methodologically, my dissertation contributes to the nascent literature on Rasch/Guttman Scenario scales. These scales are relatively new and have been employed to measure a small number of constructs (e.g., Ludlow et al., (2014) for engagement in later life activities; Chang et al., (2019) for teaching practices for equity). Students' perceptions of faculty availability present challenges as a construct that were not apparent in those applications and so have not been addressed in existing literature. For example, previous applications of this method have involved constructs where individuals' responses reflect their ratings of themselves, not perceptions of others. I also present an explicit theoretical rationale for the combination of Guttman facet design and Rasch measurement. While this may have been implicit in earlier work, it has never before been formally articulated.

Chapter Summary

This chapter introduced out-of-class communication, establishing its importance as a topic worthy of study. I also highlighted the limitations of existing instruments related to OCC with respect to construct definition and methodological approaches. I address these issues through the development of a Rasch/Guttman Scenario scale to measure students' perceptions that faculty are available outside of class. Both my construct selection and methodological process serve to enhance the validity of my instrument and ensure that its scores have clear and meaningful interpretations. In addition to elaborating on these points, this chapter also provided an overview of my dissertation's unique substantive and methodological contributions to the fields of higher education and scale development.

CHAPTER TWO: LITERATURE REVIEW

Three areas of research inform my dissertation and are reviewed here. I begin by discussing work on student-faculty interaction outside of class. This section establishes the importance of out-of-class communication (OCC) as a research area, providing a solid rationale for the pursuit of an instrument that effectively measures students' perceptions of faculty availability outside of class. Second, I review and critique existing instruments related to OCC in terms of their construct definition and developmental approaches. This section establishes the gap that my dissertation instrument seeks to fill. Finally, methodological literature related to Guttman facet theory and Rasch measurement is presented, providing the methodological background and statistical detail underlying the methodology laid out in Chapter 3. Within this section I also briefly review existing published instruments that have been developed within the Rasch/Guttman Scenario (RGS) framework and highlight what is unique about my particular application. Each section of the literature review ends with a short summary of key takeaways relevant to my research questions:

1. Can Rasch/Guttman Scenario scales provide valid and reliable measurement of student perceptions of faculty availability outside of class?
2. What is the relationship between scores from these scales and students' participation in out-of-class communication?

Student-Faculty Interaction Outside of Class

My dissertation is informed by several areas of work related to OCC. This section begins broadly, reviewing literature that establishes the importance of university student-faculty interaction as an area for research. Next, the review narrows to focus explicitly on student-faculty interaction outside of class with specific attention to four topics: (a) the relationship between OCC and undergraduate student outcomes, (b) OCC that occurs electronically, (c) the generally low participation of students in OCC, and (d) student and faculty characteristics related to OCC. These four areas come together to provide a detailed picture of the current research landscape related to OCC. They also set the stage for the second component of this literature review, where existing measures related to OCC are discussed in detail. Although these measures were instrumental in shaping the findings reported in this section, they will not be reviewed in detail here. Rather, this section focuses on establishing the importance of OCC.

The Importance of Student-Faculty Interaction

It would be difficult to overstate the importance of faculty in shaping students' postsecondary experiences. Pascarella's (1985) causal model for learning and cognitive development in college students is unequivocal about the importance of faculty as both agents of socialization and major contributors to institutional environment. Faculty contribute to undergraduate students' learning and cognitive development through both of these roles. Astin (1993) cites faculty as the second-most significant contributor to undergraduate development, eclipsed only by students' peer groups. Kuh (1995) also found that students themselves attributed gains in both interpersonal competence and academic knowledge and skills to contact with faculty.

The importance of faculty to undergraduate experiences is also emphasized in more recent work. Kuh et al. (2005) name interaction between students and faculty as a distinguishing feature of highly successful undergraduate institutions, even after accounting for various institutional factors such as size. While student-faculty interaction is considered essential across institutions, the locus of responsibility for initiating such interaction may vary across contexts. Institution size plays a key role here, with smaller institutions placing the onus of initiating contact upon faculty members, while larger institutions tend to place this responsibility upon students. Emphasis on building relationships between students and faculty is clear across successful institutions of all size, regardless of the party expected to initiate interaction. Prefacing the second topic in this section, Kuh et al. (2005) also report that not all student-faculty interaction at highly successful universities takes place face-to-face; communication via email and other electronic platforms is also essential and expected.

Chambliss and Takacs (2014) make a compelling argument concerning the centrality of faculty in successful undergraduate education. They found that interpersonal relationships in general are of utmost importance in every stage of the college-going process for 21st century students. Faculty are an important source of these relationships in multifaceted ways. While the most common-sense function of college faculty is to educate students in their areas of expertise, they play a much broader role from the student perspective:

> Students don't see disciplines in the way faculty do, as elaborate divisions of expert intellectual labor. They see instead the specific living human being standing in front of them, and from that overwhelming face-to-face reality they extrapolate the entire discipline. Because of that impact, a good teacher can redirect a student's entire college career. (p. 50)

Because students are not experts in academic disciplines when they enter college, the faculty they encounter serve as ambassadors for their respective fields. How faculty shape students' experiences and impressions of their content areas ultimately has less to do with their scholarly expertise than with their interpersonal engagement and interaction. Chambliss and Takacs (2014) report that students describe their best college professors not only as competent in their fields, but also "accessible—easy to find, available, and approachable" and "engaging" (p. 47).

Impacts of Out-of-Class Communication

A substantial body of research examines relationships between OCC and other variables of interest. Interpretation of these relationships requires knowledge of how exactly researchers conceptualized and measured OCC, which can vary considerably across the literature (see Chapter 1). These definitions and measures are not discussed at length here; they will be reviewed in the following section of this literature review. This section focuses on the results of studies establishing relationships between the frequency of OCC (however defined) and various student outcomes. Building upon the previous section that highlighted the pivotal role of student-faculty interaction in the undergraduate experience, these results establish the specific importance of OCC and its contributions to various student outcomes. Areas of ambiguity within this area of research are also discussed.

General Impacts of OCC

In a literature synthesis of faculty characteristics, Feldman (1997) found that instructor's "availability and helpfulness", including actions such as holding regular office hours and promptly responding to email were of "moderate importance" for explaining variation in students' academic outcomes (p. 104). Other studies that address specific postsecondary outcomes support this assertion. In some of the earliest OCC research, Pascarella and Terenzini (1977) found that OCC for six different purposes effectively discriminated between freshman students who did and did not re-enroll for their second year of college. These purposes were: (a) obtaining routine information related to academics, (b) discussing career plans, (c) discussing personal problems, (d) discussing academic or intellectual issues, (e) discussing campus issues, and (f) socializing. While several purposes in this list are clearly academic, others are not. These results suggest that OCC can serve more than simply an academic function.

Building on this finding, Pascarella and Terenzini (1978) report that various forms of OCC are associated with students' academic performance, as well as personal and intellectual development. Endo and Harpel (1982) examined the effects of OCC on several groups of student outcomes: personal/social outcomes, intellectual outcomes, academic achievement, and satisfaction with education. They found that OCC generally had positive impacts on these outcomes, even after accounting for student background variables. Jaasma and Koper (1999) report positive associations between OCC and student motivation, as well as students' perceptions of faculty verbal immediacy, nonverbal immediacy, and students' trust of faculty. More recently, Komarraju et al. (2010) found that OCC was associated positively with academic self-concept and intrinsic motivation. Reynolds and Ludlow (in press) found a statistically significant relationship between students' ratings of a single faculty member's availability outside of class and their overall ratings of instructor quality. Together, these findings again suggest that engagement in OCC is beneficial for students in multiple ways.

Some studies have examined OCC within the context of a particular course or content area. Looking specifically at political science courses, Guererro and Rod (2013) found that time students spent in office hours was associated with statistically significant positive changes in final course grades. Micari and Pazos (2012) examined the association between student-faculty relationships (of which OCC was a defined as a key component) and performance in a particularly challenging course: organic chemistry. They found that stronger student-faculty relationships were a statistically significant positive predictor of the students' final grades. Bjorklund et al. (2004) found student-faculty contact outside of class to be a statistically significant positive predictor of students' group work and problem solving skills, occupational awareness, and engineering competence in a first-year engineering design course. The course-specific findings complement the more general findings presented in the previous paragraph: engaging in OCC with faculty is beneficial for students both academically and personally.

Interaction with faculty can also be impactful for students beyond their time in undergraduate education. Several studies have examined the role of OCC on students' aspirations for attending graduate school. Hanson et al. (2016) found that interaction with faculty outside of class (as well as frequency of contact with faculty in general) was a statistically significant positive predictor of students' interest in attending graduate school. When additional teaching practices were incorporated into the model, interaction outside of class was no longer a statistically significant predictor, but general frequency of contact remained significant. This suggests that OCC is one of many ways through which faculty may stimulate students' interest in particular disciplines, leading them to consider further education. Similarly, Trolian and Parker (2017) examined various measures of faculty-student interaction and their relationship with students' aspirations for graduate and professional degrees. Specifically, they considered (a) frequency of faculty interaction, (b) quality of interaction, (c) research with a faculty member, (d) personal discussions with faculty, and (e) OCC with faculty. On its own, each of these was a statistically significant predictor of students' aspirations for further study; when they were brought together, only frequency of interaction with faculty and engagement in a research project remained statistically significant. It is important to note that the different measures of student-faculty interaction that Trolian and Parker (2017) employed are not necessarily mutually exclusive. For example, engaging in a research project with a faculty member likely involves OCC. It may be that the multicollinearity between these measures led to the lack of statistical significance for OCC in the full statistical model.

These findings regarding graduate school aspirations are an example of a common phenomenon within the OCC literature. While out-of-class interaction examined in isolation frequently achieves statistical significance as a positive predictor of various student outcomes, this significance may disappear when other variables are considered. For example, Umbach and Wawrzynski (2005) examined a host of faculty-related variables and their relationships with students' engagement and learning. On its own, interaction with faculty outside of class was a statistically significant predictor of engagement and learning; however, when institutional characteristics were incorporated into the model, it was no longer statistically significant. In a similar vein, Loes et al. (2012) found that a measure of faculty support (which included items related to faculty willingness to engage in OCC and faculty interest in students' nonacademic development) had only a chance outcome with measures related to students' proclivity for lifelong learning when used in a model with other predictors.

The vast majority of research related to OCC is observational. Clark et al. (2002) make the only known attempt at an experimental study concerning OCC. This dearth of experimental studies is likely because OCC is not a “treatment” that can be randomly assigned to students. Clark et al. (2002) attempted to approximate an OCC treatment by randomly assigning half of all students enrolled in a particular course to have a mandatory mid-semester meeting with their graduate student instructors for a particular course. They hypothesized that this might positively affect students’ perceptions of instructor nonverbal immediacy, affective learning, and cognitive learning. They also hypothesized that students in the “treatment” group might have higher levels of OCC, even after accounting for their mandatory mid-semester meeting. Clark et al. (2002) found that students’ perceptions of their own affective learning were significantly higher in the treatment group, and that students in this group did tend to meet with their instructors for longer periods of time, if not necessarily more often. The authors acknowledge that they employed a relatively narrow definition of OCC:

> It would be interesting to see what effect other forms of OCC might have on the frequency of office visitations. It may be the case that a casual chat after class does more to encourage such contact than a required office meeting… A possible defect in this study is that the frequency and length of other forms of OCC were not measured. (p. 835)

Student Group-Specific Impacts of Out-of-Class Communication

In addition to the broad findings described in the previous section, some research has also focused on examining differences in the effects of OCC for various student groups. A diverse array of student groups is represented within this literature, ranging from commonly-researched demographic variables such as race or gender (e.g., Anaya & Cole, 2001; Lundberg & Schreiner, 2004; Sax, et al., 2005) to academic major (e.g., Kim & Sax, 2011; Kim & Sax, 2014) to researcher-created student typologies (e.g., Kuh et al., 2000). These findings are reviewed in the following paragraphs.

Several studies address OCC with specific attention to student racial groups. This is an important area of research, as faculty play a critical role in campus racial climate and may (sometimes inadvertently) be a source of racial microaggressions for minority students (Solorzano et al., 2000). Lundberg and Schreiner (2004) examined variation in how OCC might predict student learning across different racial groups. They found that experiences with faculty explained more variation in student learning than other factors (such as student gender or institutional selectivity) across all seven racial groups included in the study. This complements the findings presented in the previous sections, which highlighted the importance of faculty-student interaction to students in the aggregate. Focusing specifically on LatinX students, Anaya and Cole (2001) found that many forms of interaction with faculty outside of class (including obtaining academic information, discussing career plans, and discussing personal problems) were statistically significant predictors of student grades, even after accounting for other variables such as class year (e.g., freshman, sophomore) and gender. This again complements other research findings.

Cole (2010) illustrates how treating minority students as a single group without attention to the nuances of racial diversity may change relationships between OCC and other variables. The author examined the effects of faculty-student interaction for minority student groups both in the aggregate and separately. When racial minority groups were aggregated, class-related contact, advice and criticism, and mentoring relationships were all statistically significant predictors of student grades; however, when the data were disaggregated into specific racial groups (Asian, African American, and Latina/o), these relationships changed. No form of OCC was a statistically significant predictor of grades for Latina/o students. For Asian and African American students, only class-related contact remained statistically significant; however, this relationship was positive for African American students and negative for Asian students.

Glass et al. (2015) examined the impact of OCC on sense of belonging for international students. Using a qualitative constructivist interview approach, they found that many international students believed their professors contributed positively to their sense of belonging, particularly through “special attention given in one-on-one conversations before and after class” (p. 358). However, recalling Cole’s (2010) findings, it is important to be aware that any relationships observed in the aggregate may not hold true for individual groups. The impacts of OCC may differ for international students with different home countries; further research is needed to explore this issue.

Focusing specifically on gender, Sax et al. (2005) found that the impact of OCC was similar for male and female students with respect to some outcomes, including overall satisfaction with faculty contact and political engagement. However, the impact of OCC was stronger for male students with respect to other outcomes (including a self-rating of competitiveness and college GPA). Exploring the frequency of OCC, Nadler and Nadler (2001) found that male students tended to have more OCC with faculty than female students. College men and women may also differ in the nature of their interaction with faculty members. Cohen (2018) reports that women are more likely to discuss logistic or procedural matters with faculty outside of class, while men are more likely to engage in intellectual discussions. Similar to the findings regarding race, these studies focused on gender suggest that there is value in examining how OCC operates across different groups of students. Examining OCC only in the aggregate may mask important issues for particular groups of students.

BrckaLorenz et al. (2017) move beyond the binary conceptualization of gender to explore the impact of OCC on participation in “high-impact” college experiences for gender-variant students (p. 350). High-impact experiences are defined as activities that increase the odds of an individual’s overall success in college, including participation in service-learning or internships (Kuh, 2008). BrckaLorenz et al. (2017) found that faculty-student interaction was a statistically significant positive predictor of the number of high-impact experiences in which gender-variant students participated. These findings suggest that OCC may be particularly important for groups of students that may struggle to find their place on campus.

Bringing many different demographic variables together, Kim and Sax (2007) looked at the relationship between three OCC-related activities (responsiveness to high faculty expectations, research with faculty for course credit, and research with faculty as a volunteer) and various student outcomes across demographic categories of gender, race, and socio-economic status. Each of these activities implies engagement in at least some OCC. Responsiveness to high faculty expectations and research with faculty for course credit were generally impactful on campus integration and gains in critical thinking and social awareness across the different student groups.

Focusing on student variation at the level of the academic unit, Kim and Sax (2011) investigated the relationship between OCC and student cognitive skills across academic majors. They found that there was a positive effect for OCC on students' cognitive skills across majors; however, these relationships were stronger in majors with "higher levels of positive faculty support" (p. 606). In a follow-up study, Kim and Sax (2014) examine the effects of OCC on academic self-concept across different academic majors, reporting that students majoring in artistic fields were more likely to both engage in OCC and be satisfied with OCC compared to other areas. The authors also reported that the strength of the relationship between OCC and academic self-concept was stronger in some academic departments than others. Looking at the effects of enrollment status on OCC, Laird and Cruce (2009) found that part-time students tended to interact less often with faculty outside of class than full-time students; in addition, higher part-time student enrollment at the institutional level was associated with less OCC for all students at the institution, regardless of individual enrollment status.

Finally, moving beyond universally-known classifications, Kuh et al. (2000) created a typology of students based on the activities in which they were most likely to engage during college; one of these activities was interacting with faculty outside of the classroom. Of the ten student types identified, the two most likely to have substantial OCC were "Artists" and "Intellectuals" (p. 238). The authors note that these findings are not surprising, as artistic fields tend to involve extended collaborations between students and faculty, while intellectually-oriented students are more likely to seek out faculty for OCC.

Out-of-Class Communication Using Electronic Media

The earliest research regarding faculty-student interaction does not include electronic communication, as the technology either did not yet exist or was not widely available. Technological advances have opened the door to a whole new realm of OCC within the past few decades. In one of the earliest examinations of student-faculty interaction via email, Atamian and DeMoville (1998) found that students were generally satisfied with their instructor's availability when all OCC was restricted to occur via email only. More recently, email responsiveness has become an important factor in determining students' satisfaction with faculty members (Sinclaire, 2014). Sheer and Fung (2007) found that many variables related to faculty email use (such as helpfulness or promptness) were positively associated with faculty-student relationship elements like trust and satisfaction.

Zhao et al. (2012) report that students frequently utilize email to contact their professors and that the content of these emails is typically related to course matters. Email is also a particularly effective form of OCC for some groups of students. For example, email is the preferred form of communication for students identifying as reserved or reticent, mainly because of comfort and ease (Kelly et al., 2001; Kelly, et al., 2004). However, there are other groups of students for whom email is viewed as a "last resort" form of OCC. Rowan-Kenyon et al. (2018) found this to be the case for first-generation college students, who generally preferred to interact with faculty face-to-face when possible.

Email responsiveness appears to be an area where students and faculty members have different expectations. Foral et al. (2010), studying both on-campus and distance students, found that students expected much shorter response times for emails than faculty believed were necessary. Additionally, faculty tended to rate themselves as more accessible via email than their students did. Faculty and students may also differ in what they consider appropriate tone for email messages. Stephens et al. (2009) found that overly casual language in emails from students to faculty negatively impacted faculty members' positive perceptions of the student as well as the student's credibility.

A few studies have also examined OCC via electronic means other than email, including social media platforms and virtual office hours. Sarapin and Morris (2015) investigated student-faculty interaction via the Facebook social media platform and found that faculty expected their relationships with students would improve through Facebook interaction. This study did not explicitly examine the student point of view; however, Rowan-Kenyon et al. (2018) found that first-generation college students preferred not to communicate with faculty through social media. Li and Pitts (2009) explored the impacts of replacing traditional in-person office hours with online office hours and found that students were generally more satisfied with a course where the option for online office hours was offered, even if they were not particularly likely to utilize them.

Lack of Out-of-Class Communication

Some research shows that students engage in OCC relatively infrequently. Cotten and Wilson (2006) assert that interacting with faculty is not a regular aspect of the university experience for many students. Fusani (1994) notes that 23% of students in his surveyed sample of 282 community college students had never interacted with an instructor outside of class and 50% of the sample did so only once or twice per semester. Despite the small sample size and particular context, Fusani's percentages are frequently cited in extant literature as evidence for an overall lack of OCC. More recently, Denzine and Pulos (2000) report that OCC is not a "spontaneous event" for traditional undergraduate students, and that students frequently discuss specific occasions for OCC with their parents or peers before reaching out to faculty (p. 58). Kuh and Hu (2001) echo this sentiment, noting that most student-faculty interaction outside of class involves obtaining specific information related to academics and that students have “relatively little personal or social contact with faculty out-of-class, such as coming together over cokes and snacks or discussing personal problems” (p. 317). These findings are an interesting contrast to the previous section, which detailed research concerning how both academic and personal OCC can have positive benefits for students.

Cotten and Wilson’s (2006) study is the most in-depth investigation of why undergraduate students do not seek out OCC. The authors employed a qualitative approach, speaking with students in nine different focus groups about their attitudes towards interacting with faculty outside of class, as well as why students did or did not engage in these kinds of interactions. Based on their focus group data, the authors propose that students themselves, faculty behaviors, and campus structure all may negatively impact participation in OCC.

Cotten and Wilson (2006) report that students are frequently unsure of whether or not faculty are open to interaction outside of class, making them reluctant to seek out OCC. When they did pursue OCC, students preferred to stick to interactions revolving around a specific problem or question regarding a course rather than engaging in free-flowing discussion; this is consistent with the findings of Kuh and Hu’s (2001). Faculty behaviors may also contribute to students’ insecurity about engaging in OCC. The authors note that students “perceive that at times faculty often behave abruptly, and they attribute this behavior most often, not to faculty time constraints, but to a lack of interest in interacting on the faculty member’s side” (p. 504). In response to these perceptions, students avoid seeking OCC so as not to bother faculty or risk their ire. Lastly, the structure of an institution's campus can also negatively influence OCC. Although students and faculty share the campus environment, they often are not present in the same physical spaces for extensive periods of time outside of class meetings.

Some research also indicates that specific groups of students may face unique challenges when it comes to OCC. For example, Anaya and Cole (2001) reported that many LatinX students perceived faculty members as either neutral or remote and unsympathetic, which may negatively impact their proclivity to seek OCC. As with the effects of OCC, reasons for not seeking out OCC may vary across different groups of students.

Examining how socio-economic status affects student-faculty interaction at elite universities, Thiele (2016) found that middle- and lower-SES students' dispositions towards OCC were markedly different from their upper-class peers. Based upon a series of qualitative interviews, Thiele (2016) placed students into three categories of interaction: "appreciative ease", "hesitant appreciation", and "critical suspicion" (p. 341). Students in the appreciative ease category were extremely comfortable interacting with faculty inside and outside the classroom. They perceived that professors were always available to them for a variety of purposes and sometimes these perceptions were at odds with faculty members' actual availability. One student that Thiele placed in this category mentioned that every professor she had ever met was "more than willing to sit down with you for like an hour and talk"—a perception that Thiele notes is likely not reflective of reality (p. 344). Unsurprisingly, upper-class students were most likely to fall into this category. Students in the hesitant appreciation category expressed nervousness about

OCC and were most likely to interact with faculty only during class or mandatory meetings. Unlike appreciative ease students, who discussed a wide range of topics with faculty, hesitant appreciation students' interactions with professors tended to stick to topics of direct relation to the course. Lower-SES students were most likely to fall into this category. Finally, students in the critical suspicion category "felt their professors were neither interested in them nor available to them as resources" (p. 347). These students were the least likely to engage in any form of OCC, and tended to be very critical of the faculty as a whole.

Jack (2016) also examines how socio-economic status may shape students' interactions with faculty and other university authority figures. Seeking to diversify the ways in which researchers conceptualize low-income students, Jack (2016) conducted a qualitative study of three groups of Black and Latino students: middle-class students, low-income students who attended boarding or preparatory high schools (called the "privileged poor"), and low-income students who attended local high schools (called the "doubly disadvantaged") (p. 5). Results showed that students in the privileged poor category, like their middle class counterparts, were more likely to feel comfortable connecting with faculty outside of class, either in-person or via email. Students in the doubly disadvantaged category not only felt uncomfortable contacting faculty, but sometimes actively avoided doing so in order to avoid being "kiss-asses" (p. 9).

Out-of-Class Communication and Student or Instructor Characteristics

Snow (1973) is the earliest study to examine faculty characteristics associated with OCC. The author instructed a sample of professors at a single institution to keep a journal of their interactions with students over the course of the semester; a subset of these individuals was also interviewed. Upper-division students at the university were then contacted by telephone and asked which, if any, faculty members they interacted with outside of class. Based on the results of this telephone survey, faculty were classified into "high", "medium", and "low" contact groups. Faculty in the high contact group tended to be untenured and did not have additional roles within their respective departments (such as part-time dean or department chair).

More recently, OCC has become a topic of interest within the communications field. In particular, many studies published in communications journals examine how various personality traits or characteristics of students and faculty can predict the frequency of OCC. Findings from these studies include the following:

- Student argumentativeness and assertiveness is positively associated with OCC (Mansson et al., 2012).
- Student communication apprehension is negatively associated with OCC (Martin & Myers, 2006).
- Instructor responsiveness is positively associated with OCC (Aylor & Oppliger, 2003).
- Instructor trustworthiness is positively associated with OCC (Nadler & Nadler, 2001).
- Student experience of "helicopter-parenting" is negatively associated with OCC (Miller-Ott, 2016).
- Instructor rapport is positively associated with OCC (Sidelinger et al., 2015).
- Instructor use of affinity-seeking strategies is positively associated with OCC (Myers et al., 2005).

- Instructor use of humor in the classroom is positively associated with OCC (Goodboy et al., 2015).

These relationships are not necessarily surprising, and some of them reflect the student and instructor determinants of OCC identified by Cotten and Wilson (2006). Students with communication apprehension or those growing up with “helicopter parents” may struggle to determine whether or not faculty are interested in interacting with them outside of class either because of personal anxiety or a lack of understanding regarding professional interaction. The instructor characteristics above that have positive associations with OCC paint a picture of a faculty member who actively engages students during class and seeks to form personal relationships with them. Such faculty members likely appear much more available for OCC than their less engaged colleagues, perhaps leading students to seek OCC at higher rates.

Section Summary

This section has established the importance of OCC and begins building a case for constructing an RGS scale related to this topic. Relationships with faculty are a central tenet of the undergraduate experience, and interacting with faculty outside of class is associated with many positive undergraduate student outcomes. Despite this, engagement in OCC is not routine for many students. It is important to acknowledge the role of socio-cultural variables in both of these findings; what is true for one group of students may not hold true for all students. Finally, relationships have been established between the frequency of OCC and various interpersonal characteristics. An RGS scale measuring students’ perceptions of faculty availability outside of class would allow for a more nuanced understanding of these relationships, perhaps enabling the development of more effective interventions leading to increased OCC for diverse groups of undergraduate students.

Existing Instruments Related to Out-of-Class Communication

Clark et al. (2002) note that OCC is "not a mainstream research area" compared to other higher education topics, such as retention or academic achievement (p. 825). This may explain the lack of a universally-accepted measure related to the construct. While there are some instruments used in multiple studies, many authors who examine OCC construct their own individualized instruments. This section of my literature review synthesizes the diverse set of measures related to OCC that are used in existing research (including many of the studies reviewed in the previous section). For organizational purposes, I classify these instruments into five groups:

1. Popular Measures of OCC: Measures in this category are used in contexts other than the study for which they were originally developed. Some instruments in this category have multiple versions.
2. Measures Extracted from Larger Instruments: Measures in this category are taken from larger survey instruments that capture the undergraduate experience more generally. These include measures from both institution-specific surveys and or nation-wide surveys such as the College Student Experiences Questionnaire (CSEQ) and National Survey of Student Engagement (NSSE).
3. Measures without a Scale: Measures in this category did not report any sort of "score" derived from multiple items to represent OCC. Rather, individual items themselves were of interest and served to capture OCC or related constructs.

4. Measures Focusing Exclusively on Email: Measures in this category focused exclusively on student-faculty communication via email and did not consider in-person interaction.
5. Miscellaneous Measures: Measures in this category were those that addressed OCC (or important OCC-related constructs), but did not necessarily fall into one of the other four categories listed above.

The following sections discuss each of these groups in detail.

Group One: Popular Measures of OCC

Knapp and Martin's (2002) Out of Class Interaction Scale is the most popular measure of OCC in extant literature. This particular measure was developed in the communications field, and this is also where it has been most widely used. There are two versions of this scale—one with 13 items and one with 9 items. Use of the 9-item scale is more common. Reliabilities (when reported) for the scale are typically quite high, ranging from .8 (Myers et al., 2006) to .88 (Mansson et al., 2012).

Mansson et al. (2012) provide a footnote listing the items in the 9-item version of the Out of Class Interaction Scale. All item responses are on a 5-point Likert scale, with 1 corresponding to strongly disagree and 5 corresponding to strongly agree. Some items are reverse scored. The items are (p. 240):

- I often talk to my instructors during their office hours.
- If I see my instructors on campus, I often talk to them.
- I rarely talk to my instructors outside of the classroom.
- If I see my instructors in the hallway, I often stop to talk to them.
- I only talk to my instructors outside of the classroom once in a while.

- I frequently talk to my instructors outside of the classroom.
- When I see my instructors around town, I usually spend some time talking to them.
- When I see my instructors in public, I avoid talking to them.
- I never talk to my instructors outside of the classroom.

All nine items relate to the frequency with which students interact with their instructors outside of the classroom. As noted in Chapter 1, this focus on frequency provides an incomplete picture of OCC. Scores from this instrument cannot reveal why OCC occurs (or not). Additionally, this instrument is restricted in its definition of OCC, only accounting for interaction that occurs face-to-face.

Another limitation of this instrument is its lack of interpretability. Respondents are prompted to think about their "instructors" in the aggregate, not to consider a specific faculty member. A score from this instrument could not be used to evaluate a particular instructor or inform them about how they might make themselves more available to students.

Knapp (2005) adapted the Out of Class Interaction Scale into the Out of Class Communication Scale for use at the level of individual faculty members. Noting the limitations of the Out of Class Interaction Scale with respect to focusing solely on frequency, Knapp (2005) added several items that address other aspects of OCC. Again, all items have 5-point Likert scale response categories. In addition to the 9 items listed above (adapted to apply to a single instructor), the Out of Class Communication Scale includes the following (again, some of which are reverse scored):

- I feel comfortable talking to my instructor outside of the classroom.

- I usually feel good after I talk to my instructor outside of the classroom.
- I enjoy talking one-on-one with my instructor outside of the classroom.
- I only talk to my instructor outside of the classroom about my grades or course material.
- I talk to my instructor outside of the classroom about topics that are not class-related (e.g., sports, movies).
- After I talk to my instructor outside of class, I like them more.
- I would like to get to know my instructor better.
- I talk to my instructor outside of class about myself and my life.
- My instructor enjoys talking to me outside of the classroom.
- It is helpful to talk to my instructor outside of the classroom.
- It is important to me to be able to talk to my instructor outside of the classroom.
- Meeting with my instructor outside of class should be a requirement.
- It is important for my instructor to know my name.
- I would like my instructor to know me better.
- If I have a problem with class content, I talk to my instructor.
- There is no reason for me to talk to my instructor outside of the classroom.
- I talk about non-class problems I have with my instructor.

Knapp (2005) performed a factor analysis, extracting four factors from these items. Unsurprisingly, one of these factors was labelled as frequency of OCC. The other three factors were labelled as attitude towards OCC, context of OCC, and topic discussed in OCC. These four factors were assigned scores and treated as separate measures for additional analyses in the study. This scale does paint a more complete picture than the original Out of Class Interaction Scale and may be used to think about individual faculty. However, its scores still suffer from a lack of interpretability because of the developmental approach applied. A score from this scale (or any of its subscale factors) would not necessarily provide actionable information for faculty or other institutional staff because the score would not be accompanied by a qualitative interpretation.

Another popular instrument originally constructed by Nadler and Nadler (2000) approached OCC from the faculty perspective. The purpose of this measure was to gauge faculty perceptions regarding quality of, satisfaction with, faculty value of, and student value of OCC. Most relevant to the present research are three 7-point Likert items asking faculty to indicate how likely students were to seek them out to discuss class matters, general academic matters, or personal problems. This scale was accompanied by a frequency report, where faculty respondents estimated the number of times per week they met with students, the average length of those meetings, the time they set aside per week for OCC, and the time actually spent per week on OCC.

Nadler and Nadler's (2000) faculty measure is complemented by a student measure regarding OCC (Nadler & Nadler, 2001). This is also a 7-point Likert scale (Cronbach's alpha = .92), where students are asked to report the likelihood of the following (p. 250):

- Recommending the instructor to a friend
- Recommending the instructor as an adviser to a friend
- Seeing the instructor outside of class about class-related matters
- Seeing the instructor outside of class about general academic matters
- Seeing the instructor outside of class about personal matters

- Taking another class with the instructor

This scale touches upon some topics that might be considered outside of the realm of OCC, such as taking another class with the instructor. A modified version of this scale is used by Dobransky and Frymier (2004), although a much lower reliability (.79) was obtained. Similar to the Knapp (2005) scale, Nadler and Nadler's (2001) scale also has limitations with respect to utility. Scores from this instrument are not clearly linked to qualitative descriptions of any time, making them difficult to interpret.

Cox et al. (2010), like Nadler and Nadler (2000), surveyed faculty to develop a measure of OCC. Items on their scales sought to capture both the frequency and content of OCC. Faculty respondents were asked to estimate how many times per week they did the following with first-year students outside of class: "discussed non-academic topics of mutual interest", "had casual conversations", "exchanged brief greetings", "discussed matters related to the student's future career", "discussed a student's personal matters," and "discussed intellectual or academic concerns" (p. 773). The authors used principal components analysis to analyze the items, ultimately extracting two components. These were named Casual Interaction (alpha = .905) and Substantive Interaction (alpha = .789). The distinguishing characteristic between these components is the depth of the OCC; this is different from other OCC instruments. For example, in other studies items related to discussing academics and personal matters tended to load on separate topical components or factors (depending upon the particular data reduction procedure employed), whereas in this measure they are both part of the Substantive Interaction subscale. However, this instrument does share one commonality with those previously described: its scores are not accompanied by clear interpretations because of the developmental and measurement approaches employed in its construction.

Wilson et al. (1974) took a different approach to measuring OCC from the faculty perspective. The authors provided the following prompt to faculty respondents:

> Faculty members have a variety of contacts with students outside the classroom. Please try to estimate how many times during the past two weeks you have met with students in the following capacities. Count only conversations of 10-15 minutes or more.

- Educational advisor – to give a student basic information and advice about his academic program;
- Career advisor – to help a student consider matters related to his future career;
- Counselor – to help a student resolve a disturbing personal problem;
- Instructor – to discuss intellectual or academic matters with a student;
- Campus citizen – to discuss a campus issue or problem with a student;
- Friend – to socialize informally with a student (p. 76).

Response categories for these items were all frequencies, ranging from “none” to “5 or more”.

Faranda (2015) used the Wilson et al. (1974) measure to inform the construction of another OCC scale. Noting that the Jaasma and Koper (1999) measure is quite lengthy to administer because of the number of items, and that the Knapp and Martin (2002) scale is very narrow in focus, Faranda (2015) created a set of fourteen 7-point Likert items (1 = strongly disagree, 7 = strongly agree) crafted to capitalize on four of the instructor roles proposed by Wilson et al. (1974): Career Advisor, Counselor, Instructor, and Friend (p.

87). Similar to many of the other measures discussed here, a principal components analysis was conducted to extract two components—one related to OCC for success in a specific course, and one related to OCC for personal/social/career reasons. However, once again, substantive interpretations cannot be made from these scale scores.

Group Two: Measures Extracted from Larger Surveys

Some studies have extracted measures of OCC from larger surveys. These surveys may be institution-specific or part of large-scale higher education data collection efforts. Surveys in this category prompt students to think about OCC generally, rather than focusing on interactions with a particular faculty member. Thus, none of them are particularly useful for providing individual-level faculty feedback. Their utility lies in their ability to provide broader pictures of the relationship between OCC engagement and other variables for large groups of students.

Using a graduating student survey created at the institutional level, Endo and Harpel (1982) extracted four factors related to OCC, naming them: (a) frequency of formal student-faculty interaction, (b) frequency of informal student-faculty interaction, (c) helpfulness of faculty members, and (d) quality of faculty advising. Formal interaction encompassed activities such as giving career advice, while informal interaction included providing extra help with coursework. Interestingly, "Number of instructors who gave academic advice" loaded strongly on both the formal and informal student-faculty interaction factors.

Some research uses OCC-related items from large-scale university surveys. For example, Kuh and Hu (2001) used the Experiences with Faculty Activity Scale from the

College Student Experiences Questionnaire (CSEQ) as a measure of OCC. The authors note that this scale was constructed to be

> Guttman-like… that is, each scale begins with items that presume less effort from students (e.g., casual conversation with a faculty member, asking about course-related matters) with subsequent items requiring more substantive exchanges and effort (e.g., seeking feedback about academic performance, worked on a research project. (p. 312)

Three components were extracted and named from these items: (a) Substantive Academic or Career-Related Interaction, (b) Out-of-Class Personal or Social Contact, and (c) Writing Improvement. The CSEQ measure was also used by Lundberg and Schreiner (2004). Mara and Mara (2011) adopted a similar approach using the National Survey of Student Engagement (NSSE); however, these authors report that the NSSE items relating to OCC were unable to capture OCC as they saw it operating in their specific context (a faculty-in-residence program).

A clear strength of the CSEQ scale is its hierarchical design, with attempts to conceive of OCC content as a continuum along which students might score from low to high. This construction may address some of the interpretational difficulties of other instruments; if a student endorses a difficult item, they should also endorse all easier items (Guttman, 1944). However, this deterministic structure rarely holds up in practice. The application of the Rasch paradigm (which is probabilistic, rather than deterministic) for scale design addresses this limitation.

Kim and Sax (2011) and Kim and Sax (2014) derived their measure of OCC from the Cooperative Institutional Research Program (CIRP) instrument. Their measures included items related to time students spent engaging in OCC, the specific nature of OCC (for example, being a guest in a professor's home), and a three-item general faculty interaction scale.

The OCC measures in this category highlight the importance of purpose in instrument construction. While these measures extracted from large surveys have utility in the context of analyses conducted with the large scale survey data, they may not be as useful on a smaller scale (e.g., for individual departments or faculty members). In their original contexts, these measures were not built to have clear, meaningful interpretations for individual respondents. It is not surprising that they fail to do so when they are used in smaller scale studies.

Group Three: Measures Without a Scale

Some studies measured OCC without representing the construct as a score derived from multiple items, instead emphasizing individual survey items as the unit of analysis. This is the approach used by Jaasma and Koper (1999), whose measure also was adapted for use in other studies. Like other instruments reviewed in previous sections, Jaasma and Koper's measure focuses on both the frequency of OCC and the content of OCC; however, they also address a third component—where and when OCC occurs. They prompted students to estimate the average number and length of interactions with faculty both "in his/her office" and "outside of class time but not in his/her office (i.e., before or after class, at break time, or on campus)" (p. 44). Students were also asked to estimate the approximate percentages of OCC time dedicated to discussing class matters, personal issues, or socializing. These estimations were followed by a single 7-point Likert item asking students to indicate their overall satisfaction with OCC. The goal was never to create a scale with an informative single score, but simply to obtain values that could be correlated with scale scores for other latent variables (in the case of the original Jaasma and Koper study, these variables were verbal immediacy, nonverbal immediacy, trust, and student motivation). Khan et al. (2015) replicated the Jaasma and Koper (1999) study with the same instrument in a non-United States context. This was likely possible because the items were treated as discrete units and not put together to form an overall scale, which may not have functioned as intended (e.g., similar reliability or factor/component structure) in a context different from that in which it was developed.

Aylor and Oppliger's (2003) measure of OCC was informed by Jaasma and Koper (1999). These authors prompted students to report frequencies for "formal" and "informal" OCC and to categorize the matters discussed during the OCC into "(1) coursework; (2) personal problems; and (3) socializing" (p. 128). Visits to office hours, email conversations, and telephone calls constituted formal OCC, while informal OCC included encounters "on campus, in the halls, at break times, at campus events, etc." (p. 128). Also like Jaasma and Koper (1999), Aylor and Oppliger's measure ended with a Likert item asking students to indicate their overall satisfaction with OCC. Again, no score was created from these items.

The items on these scales are purely descriptive. Students are prompted to report both whether or not OCC occurs, how OCC occurs, and some indication of the OCC topic (e.g., formal versus informal). However, these items cannot be interpreted as indicators of an underlying latent construct; they are simply discrete units of analysis that capture observable behaviors and actions. The power of a scale score is that it represents something that is not directly observable; this power is lacking in these measures.

The final instrument in this category is unique in that it does attend to student perceptions (a latent construct). Denzine and Pulos (2000) approached OCC in a unique way using Q-methodology. After interviewing a small sample of students about their perceptions of faculty approachability, the authors had a larger student sample complete a "Q-sort". Students were given index cards of 100 faculty characteristics and instructed to sort them into nine different categories, with category 1 representing faculty who were very unapproachable and category 9 representing faculty who were very approachable. Characteristics of approachable faculty included being willing to stay after class to meet with students, returning calls within 24 hours, and conducting one-on-one meetings with every student. Characteristics of unapproachable faculty included missing office hours, complaining about being busy, and keeping the door closed during office hours. This study is helpful in defining students' perceptions of faculty as approachable or unapproachable, but it did not yield a specific measurement instrument.

Group Four: Measures Concerning Email as a Mode of OCC

Sheer and Fung (2007) constructed an OCC scale focusing exclusively on email as the mode of communication. Their scale was composed of fifteen 5-point Likert items gauging frequency—1 was designated "very rarely" and 5 was designated "very often" (p. 293). Using principal components analysis, the authors extracted two components, which they named Professor Tasks (alpha = .78) and Social Relationship (alpha = .91). Professor Tasks included activities such as giving assignments or explaining a course policy, while Social Relationship included forwarding relevant news articles or sending holiday cards. These scales prompted students to think about the frequencies with which an instructor initiated contact using email for these purposes; it did not address the purposes for which students initiated email exchanges with faculty.

Zhao et al. (2012) adapted Sheer and Fung's (2007) scales into a four-part survey asking students about their familiarity with email, frequency with which students emailed instructors, the content of students' emails to instructors, and students' satisfaction with email as a medium for communication with instructors. They extracted two components comparable to those obtained in Sheer and Fung's original study with respect to the content of student emails to instructors—one component involved course-related tasks and one involved relationship-building.

These measures are included in this overview because of the frequent use of email for communication between instructors and students. The Zhao et al. (2012) instrument is most akin to those discussed in the previous sections. The topics covered are similar; what differs is the mode of OCC. Hence, these instruments also have similar limitations to those in other categories. Scores obtained from these scales do not have meaningful interpretations, which greatly limits their utility.

Group Five: Miscellaneous Measures

Bippus et al. (2003) asked students to report their frequency of OCC for various modes, including face-to-face conversation, email, phone, and voicemail. However, in addition to these frequencies, Bippus et al. (2003) also constructed a measure of faculty *accessibility*—a prerequisite for OCC to occur. Using twenty-six 7-point Likert items, the authors extracted two components. One of these is "social accessibility," defined as "the degree to which students view instructors as being socially available and interested in informal interaction"; the other is "physical accessibility" which is defined as "the degree to which students view instructors as being present and available for out of class interaction" (p. 263). These two facets of accessibility were hypothesized *a priori* and informed by the work of Wilson et al. (1974).

A point worth noting about the accessibility measure and its extracted components is that when an oblique rotation was used, the components were highly correlated (r = .68). Additionally, all twenty-six items had loadings greater than .3 on both the social accessibility component and the physical accessibility component. This suggests the possibility that physical and social accessibility are not necessarily distinct latent constructs, but that they might more meaningfully come together as a unitary accessibility or availability construct if a different instrument development process were used.

Another measure of OCC is Cokley et al.'s (2004) Student-Professor Interaction Scale (SPIS). This scale is not widely used and does not focus exclusively on OCC, but encompasses OCC as an important part of the student-professor relationship. Cokley et al. (2004) piloted a 73-item instrument from which nine interpretable factors were extracted, some of which only had two items. These factors were named Career Guidance, Approachable, Validity Scale (to determine student's perceptions of the importance of OCC), Caring Attitude, Off Campus Interactions, Accessibility, and Negative Experiences. Cokley et al. (2007) also validated a shortened version of these scales, retaining the same factors.

An important part of the SPIS scale worth noting is the Negative Experiences factor. Throughout the literature, there seems to be a general assumption that OCC between students and faculty is a positive occurrence—the "negative" ends of other scales indicate that students are not interacting with faculty, not that interaction is occurring with negative outcomes. Cokley et al. (2007) acknowledge the importance of not conflating quantity and quality when researching OCC:

> Although it is certainly important and worthwhile to assess the frequency of interactions between students and faculty, this methodology is limited in that it does not account for the quality of those interactions. For example, if a student meets with a faculty member 10 times during the semester for a variety of reasons, and the majority of those interactions are negative (or perceived to be negative), then, although the interactions did occur frequently, their impact is likely to be negative. (p. 54)

Fusani's (1994) measure of OCC is unique because of its response process and dual target populations. Designed for both students and faculty, respondents are instructed: "On a scale of 1 to 100, where 1 = strongly disagree and 100 = strongly agree, respond to the following questions" (p. 240). This large range of options can introduce difficulty in both selecting and interpreting a response. This instrument captures both frequency and content of OCC and reports scores for the following subscales: immediacy, satisfaction, shyness, and self-disclosure. Items for the measure were constructed based on existing instruments measuring these factors outside of the OCC context and adapted to be relevant to OCC.

One final measure of OCC is the "One-to-one" survey developed by Rosenthal et al. (2000). The exact details of how this measure was created could not be located; however, the authors situate the instrument as a way of measuring student-faculty interaction that might better inform research on student-faculty relationships:

> Their [the authors] ‘One-to-one’ survey examined ‘real interactions’ rather than ‘imagined or idealized relationships’. They argued that student-faculty ‘interactions’ are more frequent than student-faculty ‘relationships’; and that research should focus on the least positive interaction as well as the most positive a student experienced. (p. 315)

Again, it is unclear exactly what the complete form of this instrument is based upon its description; it does not appear to be used in existing literature outside of a single article. It seems to be comprised of several survey questions, along with open-response questions where students are asked to write detailed narratives of their most positive and negative experiences interacting with faculty.

The measures in this category touch upon interesting and important issues not captured in the other instrument categories. Bippus et al., (2003) explicitly attend to the importance of students’ perceptions in shaping OCC and Cokley et al. (2007) acknowledge that OCC may not always be positive. However, these and the other miscellaneous measures in this category still suffer from issues related to interpretability and meaningfulness of their scores. These issues are addressed in more detail in the following section.

Limitations of Existing Instruments Related to Out-of-Class Communication

The instruments reviewed throughout this section have enabled many valuable contributions to the literature regarding OCC. It is by using these instruments that the relationships discussed in the first section of this review were discovered. Nevertheless, they do have their limitations. These limitations can be grouped into two categories: (a) limitations related to construct definition, and (b) limitations related to measurement approaches. Each of these is reviewed in turn in the following paragraphs; this chapter then proceeds to discuss the methodological approaches and paradigms I use to address my dissertation's primary research question.

Limitations Related to Construct Definition

The measurement instruments described in the previous sections varied greatly in how their authors chose to conceptualize OCC. Some instruments (e.g., Knapp, 2005; Knapp & Martin, 2002) were chiefly concerned with the frequency of OCC between students and faculty. Many also included the topic of discussion during OCC (e.g., Aylor & Oppliger, 2003; Jaasma & Koper, 1999); however, respondents were still typically asked to address this issue from the perspective of frequency—i.e., how often do you discuss [topic] with faculty? Correlating these measures with other variables (such as student or faculty characteristics) can reveal important relationships; however, they are not necessarily useful from the perspective of providing actionable feedback to faculty or departments.

The first section of this literature reviewed discussed research related to both the benefits of OCC for students and the relative infrequency of OCC. It follows from this that faculty and university administrators have an interest in increasing OCC. Knowing how frequently students do engage in OCC (and the topics that are discussed during OCC) does not, on its own, provide actionable information for this goal. The instrument I construct measures *students' perceptions that faculty are available outside of class for engaging in OCC*, using a unique methodological approach that provides clearly interpretable scores. As will become clear in subsequent chapters, measurement of these perceptions *can* provide actionable feedback to faculty, who may then adjust their behaviors to bolster students' perceptions of availability. Of course, ultimate control over whether or not student-initiated OCC occurs remains with the students; however, finding a particular faculty member to be highly available and still experiencing low levels of OCC with students could be a useful finding in its own right. For example, it may be that while a faculty member is perceived as highly available, their in-class rapport with students is lacking, lessening students' proclivity to initiate OCC. This kind of detailed, diagnostic finding could not be obtained from any of the instruments reviewed in this chapter.

Limitations Related to Measurement Approaches

The words "factors" and "components" appeared many times throughout my review of OCC-related instruments. Colloquially, these words might be used to illustrate different aspects or parts of a particular construct; however, in the context of instrument development they imply a specific construction and set of analytical procedures. Specifically, they invoke the use of the classical test theory paradigm (CTT) (Crocker & Algina, 1986). This form of measurement, while useful in many circumstances, does not ultimately produce clearly interpretable individual scores for several reasons.

The primary analytical tools of CTT are reliability and factor analysis (Crocker & Algina, 1986; Fabrigar & Wegener, 2012). Both of these are completely dependent upon correlations among item responses. This has two implications for item development and obtaining interpretable measurement: (a) scale items are considered replications, which is what leads to high correlations among item responses, and (b) results obtained from analytical procedures are sample-dependent, as different correlations among item responses could have been obtained with a different group of respondents. This means that scales developed within the CTT paradigm cannot place individuals along a meaningful continuum of the construct of interest (meaning that they have no clear interpretation with respect to the construct), nor can respondents and items be directly compared. These limitations of CTT are addressed in this dissertation through the use of the Rasch paradigm for measurement, which is presented in the following section.

Section Summary

This section discussed existing instruments related to OCC and their limitations. There are a few key takeaways. One of these is that although OCC itself is a relatively simple concept, there are many different ways that researchers might choose to approach its measurement. Many of the instruments described in this section made use of factor analytic procedures within the classical test theory paradigm. Others employed unique approaches such as Q-methodology or the one-to-one survey. The methodological approaches limit the interpretability of scores from existing OCC-related instruments, which compromises their utility for purposes such as faculty evaluation or departmental improvement.

Methodological Approaches and Paradigms

In my dissertation, I construct a Rasch/Guttman Scenario (RGS) scale to measure students' perceptions of faculty availability outside of class. The RGS scale development process is informed by two methodological approaches: Rasch measurement and Guttman's facet theory. These two distinct bodies of literature can be brought together in a compelling way to aid in the construction of scales that are both substantively meaningful and technically sound (i.e. RGS scales). First, Rasch measurement and Guttman's facet theory are discussed individually. These presentations are followed by a synthesis of their commonalities and differences and an overview of RGS development procedures. Finally, existing instruments measuring other constructs that employ the RGS methodology are briefly reviewed in order to highlight my dissertation's unique contributions to this emerging area of scale development literature.

Rasch Measurement

The Rasch Paradigm

Although Rasch measurement (Rasch 1960/1980) is often used to refer to a set of analytical procedures, it is perhaps more appropriate to describe it as an analytical paradigm. This paradigm differs significantly from both classical test theory (CTT) (Crocker & Algina, 1986) and other item response theory (IRT) approaches (de Ayala, 2009). CTT and IRT both stress the importance of analytic approaches that appropriately *model the data*, whether it be through the extraction of factors based on item response correlations (CTT) or the inclusion of multiple item parameters during model estimation (IRT).

In contrast, Rasch measurement emphasizes the importance of invariance. Andrich (2004) describes invariance: "…briefly, that the comparison between any 2 persons should be independent of which items from a class of items is used, and vice versa. The case is not that it describes data" (p. 8). This concept has also been called specific objectivity (Rasch 1960/1980) and objective measurement (Wright, 1967). Engelhard (2013) presents five requirements necessary for achieving invariant measurement (p. 14):

1. Item-invariant measurement of persons; i.e., person ability estimates are independent of specific items encountered

2. Non-crossing person response functions; i.e., more-able individuals have greater chances of item success or endorsement than less-able individuals on all items
3. Person-invariant calibration of items; i.e., item calibrations are independent of specific individuals used for calibration
4. Non-crossing item response functions; i.e., any person must have a greater chance of success on an easier item than a harder item
5. The existence of a single underlying latent variable with a continuum along which persons and items are placed

The technical details of how these criteria are achieved are presented shortly; the purpose of this introductory discussion is to highlight the unique aspects of the Rasch measurement approach that make it particularly useful for examining latent constructs. Latent constructs are not observable and therefore are not directly measurable in and of themselves; rather, individuals' responses to items deliberately designed to activate the latent construct are used as indicators (Hambleton & Cook, 1977). Students' perceptions of faculty availability outside of class are an example of such a construct; "perceptions" are not directly observable, so scale items serve as indicators of those perceptions. Invariance is a highly desirable property of scales measuring latent constructs, as it ensures that scale scores across groups of respondents can be meaningfully compared and that respondents and items can be meaningfully compared to each other with respect to the underlying construct continuum.

Principles of Rasch Measurement

The principles of Rasch measurement are extensions of the work of Thurstone (1959), which presents one of the earliest arguments for measurement of attitudes through the conceptualization of a latent construct or trait. These principles guided the seminal work of Rasch (1960/80) and are presented concisely in Ludlow et al. (2014). Adherence to these design principles, along with the technical estimation components, results in measurement instruments that meet the requirements of invariance. As mentioned above, an essential aspect of this paradigm is the emphasis on data fitting the model, rather than the other way around (as in other item response theory approaches). Thus, if it is found that collected data do not fit the Rasch model (determined through means described below), the prescribed course of action is not to adjust the model, but rather to reconsider the construction of the instrument or how data were obtained (Andrich, 2004).

The principles of Rasch measurement, explicated in Ludlow et al. (2014), are presented in the following list. Details as to how these principles may be put into practice are discussed in subsequent paragraphs.

1. Scale items should measure a single, uniform construct.
2. Scale items should represent a range of difficulty (whatever difficulty means in the given context), with some items being easy and others being difficult.
3. The scale items are adequately spread across the difficulty continuum; there should not be large gaps in difficulty such that some levels of the construct are not captured by the items.
4. The continuum of difficulty along which the scale items fall follows a clear progression.
5. All scale items should have the same discrimination, or relationship, to the overall scale score.

6. All scale items should be independent, meaning that the way that an individual responds to one item should not influence how they respond to other items.
7. Scale items that do not seem to fit the theory underlying the model should be modified or eliminated.

The first principle stresses the importance that items included in a Rasch-based scale all represent a unidimensional latent construct. This is important because it ensures that scores on the instrument are reflective only of individual's levels of the latent trait and that variation in item responses is not due to constructs that are not of interest. Smith and Miao (1994) suggest that unidimensionality may be assessed using both fit statistics produced in Rasch analysis software and application of factor analytic techniques prior to fitting the Rasch model. Problematic Rasch fit statistics may be interpreted a myriad of ways; the authors argue that lack of unidimensionality is one relevant interpretation. Linacre (1998) advises against the use of factor analytic procedures ahead of the Rasch model for assessing unidimensionality, noting that it is possible for different "factors" to appear reflecting levels of the unidimensional construct of interest. This might lead one to incorrectly reject the assumption of unidimensionality. A final option for assessing unidimensionality is performing a principal components analysis upon the residuals obtained from a Rasch solution (Ludlow, 1985); extraction of interpretable components may serve as an indication that there is an additional dimension underlying the data (Tennant & Pallant, 2006).

Principles two through four are logical to consider together, as they all concern the continuum of the underlying construct and how scale items relate to this continuum. Scale items should be deliberately authored to reflect variation on the construct of interest. Item spread along the continuum is important because it ensures that there are items adequately capturing and describing all levels of the construct. Lastly, the ordering of items along the continuum should be reasonable and aligned with theory-driven expectations. Related to the evaluation of these principles, Wright and Masters (1982) suggest reflection upon the following questions: "Have we succeeded in defining a discernable line of increasing intensity?" "Is item placement along this line reasonable?" (p. 90). These questions may be addressed through careful examination of the variable map that is produced by Rasch analysis software. The variable map is a visualization of the latent trait continuum, along which both people and items are placed at their respective locations (Engelhard, 2013).

Principle five, equal item discrimination, is key in distinguishing the Rasch approach to modelling from other item response theory models. Equal discrimination means that the item response functions for all items comprising an instrument will have the same slope, implying that all items are assumed to correlate equally with the overall scores on the instrument. The principle again speaks to the importance of data fit to the model within the Rasch framework. In contrast, multi-parameter item response models will allow item discriminations to vary. This inevitably leads to better indices of model fit (de Ayala, 2009), but compromises several of the premises of invariance described by Engelhard (2013), namely that person- and item-response probability functions do not intersect or cross when they are plotted. Unequal or poor discriminations in items can be detected in the Rasch model using fit statistics (discussed below).

The sixth principle refers to independence of observations. This means that persons' responses to individual items do not influence their responses to other items; this is also known as the principle of local independence (Bock, 1997). It is not possible to prove that this assumption holds; however, steps can be taken at the item-writing stage to guard against item dependency. For example, items should not directly reference each other or build off of each other's answers.

Finally, the seventh principle refers to the importance of optimizing the scale to provide meaningful and invariant measurement. If specific items are found to be problematic, whether by fit indices or placement on the variable map, they should be revised or removed. This again highlights the perspective of collecting data with an eye towards fitting the model.

These principles make clear that explicit consideration of theory underlying a construct is integral to Rasch measurement. An *a priori* conception of how the construct of interest is structured along a hierarchical continuum is essential for assessing how well any data that are collected conform to the model. Returning to the questions posed by Wright and Masters (1982), it is impossible to determine whether or not a "discernable line of increasing intensity" has been created without having some theoretical idea of what a meaningful progression along the construct would be (p. 90). Similarly, the reasonableness of item placement along this line cannot be assessed without upfront consideration of what such reasonableness would look like.

Technical Details of the Rasch Model

The essential premise of the Rasch model is that individuals' responses to items are governed probabilistically by person ability and item difficulty (Wright, 1967). Ability and difficulty are the terms used within Rasch literature for individuals' and items' levels on the underlying latent trait, although this particular vocabulary may not be ideal for all contexts. As the Rasch model is not deterministic, it does not govern response patterns without error; however, generally respondents will succeed (assuming a binary outcome) on items where their ability is greater than the item's difficulty and will not succeed on items where their ability is less than the item's difficulty. The simplest form of the Rasch model, again assuming a "right/wrong" binary outcome is presented in the following equation:

$$\pi_{nix} = \frac{e^{(\beta n - \delta_{ix})}}{1 + e^{(\beta n - \delta_{ix})}} \tag{1}$$

Where:

- π_{nix} is the probability of person n selecting correct response x to item i.
- β_n is the person parameter representing person n's level (ability) on the underlying latent trait.
- δ_{ix} is the item parameter representing item i's level (difficulty) on the underlying latent trait.
- e is the exponential function constant.

According to the equation above, individual n's probability of correct response x to item i is determined by the person parameter for ability (β_n) and the item parameter for difficulty (δ_{ix}). If the person and item parameters are equal (i.e., perfectly matched), an individual's probability of a correct response is .5. As person ability becomes greater than item difficulty, the probability of a correct response increases; conversely, as person ability becomes less than item difficulty, the probability of a correct response decreases.

At this stage it is important to note the units in which person and item parameters (β_n and δ_{ix}) are expressed, as one of the defining features of invariant measurement is that people and items are both captured on the same scale (Engelhard, 2013). Wright and Masters (1982) describe an important property of measurement units: “A unit of measurement is always a process of some kind which can be repeated without modification in the different parts of the continuum” (p. 2). In the case of the Rasch model, the unit of measurement is referred to as a logit. Simply put, logits are the unit resulting from taking the log of an odds ratio (i.e., the probability of an event occurring divided by the probability of an event not occurring). Applied to persons and items, a person ability estimate in logits is the natural log odds of that individual succeeding on an item with a 0-logit difficulty, while an item difficulty estimate in logits is the natural log odds of an item being failed (i.e., incorrect response) by an individual with a 0-logit ability (Ludlow & Haley, 1995).

The use of logits allows for the creation of a scale with interval-level properties and a meaningful zero point. Difficult items and individuals with high ability will have logit estimates greater than 0, while easy items and individuals with low ability will have logit estimates less than 0. There are no official bounds for how high or low these estimates can be, although it is uncommon to see estimates greater than 6 or less than -6.

Naturally, the next question to arise concerns how these logit estimates for persons and items are obtained. Invariant measurement mandates that these be estimated separately (Engelhard, 2013). Wright and Masters (1982) refer to this concept as separability, providing the following definition:

> Separability means that the connection between observations and parameters in the measurement model can be factored so that each parameter and its associated

statistics appear as a separate multiplicative component in the modelled likelihood of a suitable set of data. (p. 8)

Rasch (1966) provides an illustration of how person and item parameters may be separated for estimation. This separability has another important implication for estimation known as sufficiency. When person and item parameters are estimated separately, the total raw scores for persons are sufficient statistics for the estimation of item parameters; likewise, total raw scores for items are sufficient statistics for the estimation of person parameters. In neither case is it necessary to know specific patterns of responses (Rasch, 1966).

The concepts of separability and sufficiency are essential in the application of Rasch models; however, they themselves are not estimation procedures for obtaining person and item parameters. The procedure most commonly used to execute these estimations is the joint maximum likelihood estimation (JMLE) method (Engelhard, 2013). Correct application of JMLE is dependent on the Rasch principle of local independence, which mandates the "statistical independence of response from one item to another of persons having the same value as that of the underlying latent variable" (Bock, 1997, p. 24). Assuming local independence, it is possible to calculate the likelihood of a particular pattern of responses for an individual using the following equation:

$$P(\beta) = \prod_{i=1}^{L} P_i^{xi}(1 - P_i)^{1-xi} \tag{2}$$

Where:

- xi is the dichotomous response of an individual for item i.

- $P_i(\beta)$ is the conditional probability of a person's success on an item according to the model (see the general Rasch model equation above).
- $\prod$ is the multiplication operator across the L items.

Knowing the item parameter estimates allows for maximum likelihood estimation of person parameters; conversely, this same equation can be applied to estimate item parameters when person parameter estimates are known (Engelhard, 2013). As these parameters are estimated simultaneously in practice, the sufficiency of person total raw scores for the estimation of item difficulty and of item total raw scores for the estimation of person ability are essential (Rasch, 1966). JMLE is an iterative process; after the generation of initial estimates for person and item parameters, the Newton-Raphson algorithm (in most Rasch software programs) is used to optimize these estimates (Ludlow & Haley, 1999). The eventual result is independent logit estimates of persons and items, enabling their meaningful comparison on the same scale.

The Rasch Rating Scale Model

The details provided in previous sections were presented in reference to the dichotomous Rasch model. Rasch models can take many different forms, each of which are most suitable for specific purposes. For scale development, a particularly useful form of the Rasch model is the rating scale model (Andrich, 1978). Unlike the dichotomous model described above, which is most appropriate for binary "correct/incorrect" items, the rating scale model is best suited for items with ordered response options—for example, a set of response options ranging from Strongly Agree to Strongly Disagree (Wright & Masters, 1982). Rather than defining difficulty in terms of being "correct", the rating scale conceptualizes difficulty with respect to moving from one response category to the next. In this case, “ability” may no longer be the most appropriate word to describe the person parameter; it may more appropriately be described as the person level on the underlying latent trait. Despite these conceptual differences, the interpretation of person and item logit estimates is comparable to the dichotomous case. Individuals with higher logit estimates have higher levels of the latent trait; correspondingly, items with higher logit estimates may represent higher latent trait levels (Andrich, 1978). When the rating scale model is used, it is usually no longer appropriate to discuss the idea of “success” on an item—it is more correct to use the notion of “endorsement” of a response category.

In addition to an average difficulty measure for the item, the rating scale model provides a set of thresholds for response options that indicate the relative difficulty of moving into a specific response category compared to the response category below it (for example, the difficulty of going from “agree” to “strongly agree”). These thresholds are held constant across all items; Wright and Masters (1982) provide the following rationale for this constraint:

> The relative difficulties of the “steps” in a rating scale item are usually intended to be governed by the fixed set of rating points that accompany the items. As the same set of points is used with every item, it is usually thought that the relative difficulties of the steps in each item should not vary from item to item. (p. 48)

The rating scale model is depicted in the following equation:

$$\pi_{nix} = \frac{e^{\sum_{j=0}^{x_{ni}}[\beta_n - (\delta_i + \tau_j)]}}{\sum_{k=0}^{m} e^{\sum_{j=0}^{x_{ni}}[\beta_n - (\delta_i + \tau_j)]}} \tag{3}$$

Where:

- π_{nix} is the probability of person *n* responding to item *i* in category *x*.

- β_n is the "ability" estimate for person *n* (Note that the definition of "ability" varies depending upon the context).
- δ_i is the estimate of difficulty for item *i.*
- τ is the threshold difficulty for the *k*th step from one response category to the next.

The only difference between the rating scale model and the dichotomous Rasch model is the inclusion of threshold estimate τ. Andrich (1978) explains that τ

> …qualifies the affective value of the item… This qualification may show for example that an agree response to an item with moderate affective value is equivalent to a neutral response to an item of high affective value. (p. 565)

These qualifications are constrained to be the same across all items in the scale, ensuring the retention of meaningful interpretation.

Evaluating a Scale with the Rasch Model

As noted in the preceding discussions, effective execution of Rasch analysis requires a detailed *a priori* conception of the latent construct of interest, including a clear articulation of what it means for individuals to progress from low to high on the construct (Ludlow et al., 2014). Again, the goal is to determine whether or not the data adequately fit the model, not the other way around (Andrich, 2004). There are several key results in a Rasch analysis that can aid in making this determination. The following paragraphs discuss these in detail.

Variable Map. An important first step in evaluating a Rasch-based scale involves examination of the variable map. The variable map is a graphical representation of persons, items, and their corresponding logit estimates (Engelhard, 2013). Items and persons with low logit estimates appear at the bottom of the map, while items and persons with high logit estimates appear at the top of the map. The variable map allows the assessment of scale performance through consideration of the following questions (Wright & Masters, 1982):

- Are the scale items in the intended order from easiest to most difficult?
- Are there large gaps in the scale where there are no items?
- Are the respondents and the items generally well-matched? Or does the scale appear to be too easy or too difficult for the respondents?

Answering these questions allows determination of whether or not the scale items are aligned to the principles of Rasch measurement described in previous sections. For example, a large gap in the variable map with no items would violate the principle of adequate spread of items across the continuum. Such a finding might indicate the need for additional items to adequately capture the level on the underlying construct where the gap occurs. If items are not arranged in the variable map in the hypothesized order, this might be suggestive of a need to revise one's understanding of the construct. This point is somewhat belabored, but it is essential to highlight that problems surfacing through examination of the variable map must be addressed through either a theoretical adjustment to one's understanding of the latent variable or modification of data collection procedures; adjustments are not made to the Rasch model (Andrich, 2004).

Fit Statistics. The process of Rasch estimation provides goodness-of-fit statistics for both people and items. As Rasch models are probabilistic, these statistics provide an indication of how well the matrix of observed person responses to items aligns to what was expected under the model (Wright & Masters, 1982). For example, the following equation provides an expected value of response x_{ni} for each person-item interaction in a data:

$$E_{ni} = \sum_{k=0}^{m} k\pi_{nik}$$

Where:

- π_{nik} is the modeled probability of person n selecting category k for item i.
- m is the number of response options.

These expected values allow for the calculation of residuals by subtracting an individual's expected response to an item from their observed response to an item. A large positive residual would be indicative of an unexpectedly high response, while a large negative residual would be indicative of an unexpectedly low response (Ludlow, 1983; Wright & Masters, 1982).

Variances of these residuals can be computed, allowing for their standardization. It is these standardized residuals that are used to calculate person and item fit statistics (Wright & Masters, 1982). Both weighted and unweighted versions of these statistics may be calculated (Engelhard, 2013; Smith & Plackner, 2009). Often referred to as "infit" (weighted) and "outfit" (unweighted) statistics in Rasch reporting, these two forms of fit statistics reveal different types of misfit. Unweighted fit statistics are useful for diagnosing the presence of individual outliers, while weighted fit statistics are more helpful in diagnosing patterns of unexpected responses (Linacre, 2002). Both the unweighted and weighted fit statistics are expressed as mean squares calculated through the use of standardized residuals (Wolfe & Smith, 2007):

$$MS_{unweighted} = \frac{\sum_{n=1}^{N} z_{ni}^2}{N}$$

(5)

$$MS_{weighted} = \frac{\sum_{n=1}^{N} z_{ni}^2 W_{ni}}{\sum_{n=1}^{N} W_{ni}}$$

(6)

Where:

- z is a standardized residual.
- W is the information or variance for each item. This ensures that items with greater variance are given greater weight.

The expected value for these statistics is 1, with greater values indicating misfit and lesser values indicating overfit. Although there are no hard cutoffs for flagging problematic unweighted and weighted mean squares, a general criterion of 1.4 has been proposed for both measures (Wolfe & Smith, 2007). Ludlow et al. (2014) propose a somewhat lower criterion of 1.2 or 1.3 for investigation during early stages of instrument development so as to avoid missing potentially problematic items or individuals.

These mean squares can also be transformed into an approximate "t" statistic. Generally, absolute values higher than 2 merit investigation; however, these statistics are heavily influenced by sample size, so a more stringent criterion is often applied (Wright & Masters, 1982).

Item/Category Characteristic Curves. Item characteristic curves (ICCs) provide a graphical representation of probabilities of individuals at different ability levels succeeding on an item (Andrich, 1988). In the rating scale case, these graphs, called category characteristic curves (CCCs), represent the probabilities of individuals at different levels on the underlying latent variable selecting different response options

(Wright & Masters, 1982). The points at which curves for different response categories intersect are the threshold difficulties (τ) described above. The following questions can be answered through examination of CCCs (Wolfe & Smith, 2007):

- Are all response options being utilized by respondents?
- Is the curve for each response option unimodal?
- Does the order of the category curves proceed in a reasonable manner?

Wolfe and Smith (2007) recommend that a minimum of 10 respondents make use of each response category. Ideally, each response option should have one (and only one) place in the latent trait distribution where it is most likely to be chosen. If this is not the case it may be defensible to collapse response categories for analysis. Additionally, given the conceptual logic of the rating scale model (Andrich, 1978), across all items, it should require a higher level of an underlying trait to endorse a "higher" response option—e.g., "strongly agreeing" with a statement should be more difficult than "agreeing" with a statement (reverse coded items excepted).

Reliability and Separation. Reliability and separation, also provided for both persons and items, provide an assessment of the degree to which the instrument is able to reliably distinguish between individuals who are high and low on the latent trait (Wright & Masters, 1982). Separation concerns either the classification of respondents (person separation) or establishment of item hierarchy (item separation). Person separation above 2 and item separation above 3 is desirable; if values are lower than these thresholds there is a risk that either (*1*) the instrument cannot effectively discriminate between individuals with high and low levels on the latent trait, or (*2*) that the sample of persons used is not sufficient to confirm the item hierarchy (Linacre, 2018). Reliability for persons in this context is comparable to traditional test reliability; it represents the proportion of "true" variance represented in a set of scores. A criterion of .8 is suggested. Reliability for items refers to the spread of items across the hypothesized continuum; a criterion of .9 is suggested (Linacre, 2018).

There are many different factors that affect reliability and separation for persons and items. For persons, these include the ability variance of the sample, the length of the test, the number of response categories, and the alignment between the sample and items. Greater ability variance, longer test length, and higher numbers of response categories all contribute to higher person reliabilities. Person reliability is also increased when the items are well-aligned with the abilities of the persons in the sample (Linacre, 2018). Item reliability depends on variance of item difficulties and number of persons in the sample; increases in both of these contribute to higher item reliabilities (Linacre, 2018).

Guttman Facet Theory

Guttman's facet theory (Guttman, 1954; Guttman, 1959) has both a design and an analysis element. The design element is a particularly useful complement to the Rasch measurement paradigm described in the previous section and is employed in the construction of RGS scales. The following sections will provide a general overview of facet theory principles, followed by a more detailed discussion of facet theory design and analysis.

Principles of Facet Theory

Both the design and analysis components of facet theory are rooted in the concept of facets. Guttman (1954) defines a facet as a "set of elements" that are useful in "designing universes of content for research projects" (p. 399). Shye (2015, p. 149) presents seven principles underlying facet theory. These principles encompass the entirety of facet theory, meaning that they address both design and analysis elements. The seven principles are listed below and the following paragraphs provide more detailed discussion.

1. Attributes examined through facet theory approaches may be represented in a geometric space.
2. Variables related to the attribute are points in that space.
3. *Every* point in the space represents a variable of the attribute.
4. Variables selected for smallest space analysis (SSA; the primary analytic technique in facet theory) must belong to the same content universe or attribute.
5. Observed variables within any given analysis are only a sample drawn from all of the variables comprising a content universe or attribute.
6. The placement of observed sample variables within the geometric space of an attribute can guide the partitioning of the space into components or sub-domains.
7. Relationships between these components are expressed geometrically rather than algebraically.

Principles three through five primarily concern facet theory design elements, while principles one, two, six, and seven address analytical elements. The most important takeaway from a design perspective is that any observed variables (item responses in the context of survey-type research) are merely a sample of the multitude of variables that might have been chosen to represent the attribute in question. Principle four, although invoking the specific analytical technique of SSA, brings to the forefront another important design consideration: in order for variables to be used together within an SSA, they must belong to the same "content universe" (Shye, 2015, p. 149). This highlights the importance of either (a) a thorough understanding of the attribute being examined, or (b) a well-developed hypothesis about the nature of the attribute in cases where no previously formally articulated theory exists.

Facet theory principles one and two serve as the conceptual foundation for the theory's preferred analytical procedures. Facet theory makes a clear distinction between two types of data analysis: those that are intrinsic and those that are extrinsic. Intrinsic data analysis is termed such because it "adheres closely to the defining features of the data" (Borg & Shye, 1995). The first two principles in the list above explicate those features in the facet theory framework. Subsequently, principles six and seven specify how these defining data features manifest in the interpretation of results from intrinsic analytical procedures.

Facet Theory Design Elements

Borg and Shye (1995) describe facet theory as a formalized way to define a given construct or domain and explain its structure. Guttman and Greenbaum (1998) note that theory has two general components: "a framework for defining the universe of observations [1], together with the empirical distribution of the observations, carried out within the design of the framework [2]" (p. 31). In this way, facet theory can inform systematic data collection about a construct of interest. Succinctly, use of facet theory forces researchers to name "concepts and contexts that guide empirical observations" for a given construct (Borg & Shye, 1995, p. 13). These concepts and contexts are the "facets" of a construct.

Naturally, a key component of facet theory is the identification of facets to represent a construct. Building upon Guttman's (1954) definition, Borg and Shye (1995) define facets in this way: "A facet is a set of elements (i.e., types, classes, categories, attributes, etc.) that classify objects of interest" (p. 25). Canter (1985) describes three general categories of facets: the common range facet, the population facet, and the content domain facet. The common range facet provides a general structure for how individuals might respond to items, the population facet defines the individuals to be studied, and the content domain facet(s) are used to describe the population's experience of the phenomena of interest. Facets are useful in that they provide a framework for systematically examining and naming important aspects of a construct, providing a means to "*structure* a given unstructured universe" (Borg & Shye, 1995 p. 32, emphasis original). This approach is particularly helpful for complex constructs that are not easily defined. However, it must always be kept in mind that any facets and their corresponding values represent only a sample of what might have been observed (Shye, 2015).

A key tool in the identification and delineation of facets for a given construct is the mapping sentence. A mapping sentence is a semantic structure within which one situates the facets of a given domain. Each facet represented within a mapping sentence is known as a *struct*, while the unique combination of facet elements represented in a given mapping sentence is known as a *structuple*. Borg and Shye (1995) provide the following example of a mapping sentence designed to capture individuals' perceptions of how quickly time passes (p. 2):

Positivity
($p1$ = pleasant)
Person (p) believes that a ($p2$ = neutral) situation with
($p3$ = unpleasant)

Number | Variability | | Difficulty
($m1$ = many) ($v1$ = variable) ($s1$ = difficult)
( ) ( ) events that are ( ) to
handle
($m2$ = few) ($v2$ = monotonous) ($s2$ = easy)

Reaction
(very short in duration)
are felt as → (to)
(very long in duration)

This example has four facets, or structs: positivity, number, variability, and difficulty. Each struct has several different possible values; it is not necessary that all structs have the same number of values, as this will depend on the nature of the facet that is represented by the struct. An example of a structuple from this example mapping sentence would be: Person (p) believes that a neutral situation with many variable events that are easy to handle would be felt as…

The mapping sentence is arguably the most important tool of facet theory. Mapping sentences provide a systematic structure within which to situate and manipulate facets, making them "natural starting points for model building" (Borg & Shye, 1995, p. 57). Hackett (2014) expounds upon this point, saying:

> The mapping sentence provides a theoretical hypothesis of how an individual 'understands' the reality of a given sector of his or her experiential life. The mapping sentence is therefore a conceptual and empirical framework that allows comparisons between people without imposing propositions of understanding upon individuals. A mapping sentence specifies the logical relationships amongst

> facets and facet elements and by so doing it also assists in the identification of redundancy in facets and their elements as well as suggesting areas where adequate description of research content is missing. (p. 46)

The principles of facet theory and the mapping sentence tool can be used to generate items for measurement purposes through the notion of the common range. Borg and Shye (1995) describe the common range as "established by substantive considerations on the *common meaning* of the ranges [of item responses] (p. 61, emphasis in original). This implies that individuals' responses to items are all indicative of the same kind of construct, for example an attitude or frequency (Canter, 1985). The authors go on to note that common ranges are rarely discovered post hoc; more often, they are conceptualized in advance and then used to inform the selection of meaningful facets. Additionally, levels of various facets should be monotonically related to the common range; without this kind of relationship, it is not possible to support the assertion that all facets point to a single construct. Borg and Shye (1995) discuss the perceived tension between the multidimensionality that is seemingly invoked by the use of facets and the idea of the common range, saying "it is a common fallacy to assume empirical 'multidimensionality' somehow disproves the hypothesis of the common range" (p. 62). Rather, the common range refers to the similarities across items intended to measure a single domain or construct, while facets represent explicit identification of how items that are presented together tap into the domain or construct.

Facet Theory Analysis Procedures

Several of the facet theory principles move beyond the use of facets for design purposes and explicitly invoke analytical procedures. Specifically, facet analysis calls for the use of procedures such as faceted smallest space analysis (FSSA) and partial order scalogram analysis by base coordinates (POSAC) (Shye, 1998). These procedures are rooted in the representation of constructs as geometric space, within which observations are situated. The ultimate outcomes of both FSSA and POSAC are graphical, hence their classification as intrinsic procedures that adhere closely to data structure defined as a geometric space. FSSA provides a division of geometric space into meaningful areas, while POSAC generates profiles of individuals based upon the observations obtained about them (Shye, 2015).

Critiques of facet theory typically concern the analytical procedures rather than the design elements. As Borg and Shye (1995) note, it is difficult to make a case against employing systematic design principles in the construction of theory or a domain. However, on the analytical side, there is one notable criticism of the intrinsic procedure of FSSA: its lack of formal acknowledgement or treatment of uncertainty (de Souza et al., 2014). de Souza et al. (2014) note that this lack of acknowledgement likely comes from the fact that facet theory does not operate within the traditional realm of statistics, instead relying on its intrinsic set of analytical procedures. The authors also discuss how some applications of FSSA generate scalograms where authors have made clear modifications, calling this "straightforward allowances for errors and random variations that are assumed (but not proven) to be of no theoretical or empirical significance" (p. 65). Finally, de Souza et al. (2014) found that scalograms generated through FSSA differed substantially when random subsamples of respondents were analyzed from an overall sample.

Combining the Rasch Paradigm and Facet Theory Design

The information presented in the previous two sections provides an overview of both Rasch measurement and facet theory. This section brings the two ways of thinking together, beginning with a theoretical justification for marrying the two theories before presenting the RGS development methodology and its existing applications.

Theoretical Justification

This section will discuss where there is a meaningful theoretical overlap between these Rasch measurement and facet design, as well as present an argument for their conjunctive use. Generally, I argue that the use of facet design (and the concepts underlying facet design) are a useful way to construct survey items in such a way as to meet the principles of Rasch measurement. Conversely, the concept of invariant measurement and probabilistic model underlying the Rasch paradigm address some of the limitations of facet theory analytical procedures.

As discussed in detail earlier, the principles of Rasch measurement generally mandate that items cover a wide range of difficulties along a single unidimensional construct of interest. Items should be well-spread along the continuum of the construct and show a clear progression from high to low (recognizing the exact vocabulary to describe this progression may vary across different kinds of constructs) (Ludlow et al., 2014). These notions are complementary to the facet design principle of the common range, which implies that all observations (items) are indicative of a single domain or construct, and mandates that levels of facets sharing a common range should increase monotonically. Bringing the two paradigms together explicitly, the common range of facet theory and the unidimensionality assumption of Rasch measurement both mandate explicit, *a priori* consideration of the construct of interest and deliberate construction of items or observations to ensure adequate coverage of that construct.

In addition to this key underlying commonality, facet theory design also provides an important tool that can be used for the crafting of items that subscribe to the common range and Rasch principles: the mapping sentence. As shown above, mapping sentences provide an organizational structure for the facets that have been selected to represent a given construct. Not only do mapping sentences name those facets, but they also contain information about the various levels of facets. The levels of included facets can be systematically varied across items, allowing the deliberate and controlled crafting of items that purposefully span a range of levels within the overall domain or construct. In summary, the mapping sentence is a useful lexical tool for the authoring of items that are likely to meet the Rasch principles related to spread along the construct continuum. Mapping sentences also provide a clear foundation for actual item text, although this may be modified to ensure that items are clear and engaging for respondents.

The analytical procedures belonging to the Rasch paradigm likely would not be considered intrinsic according to prominent facet theory authors (e.g., Borg & Shye, 1995), as they do not necessarily adhere to the conceptualization of data as points in a space. However, the Rasch paradigm does mandate close alignment between analysis and data in its own way. As mentioned above, when the Rasch model does not fit collected data, the model itself is never altered; rather, understanding of the construct is refined or data are collected under different conditions (Andrich, 2004). As such, while the exact analytic procedures of Rasch measurement may not be explicitly intrinsic within the facet theory framework, there is alignment between the overall goals of intrinsic data analysis and Rasch measurement: analysis and data should be adequately matched. Both analysis procedures also produce graphical output that may be used to meaningfully partition the construct of interest (variable maps in Rasch measurement and scalograms in facet theory analysis procedures).

Rasch measurement also addresses some of the limitations of facet theory analytical procedures through its emphasis on invariance. For example, de Souza et al. (2014) found substantial evidence that FSSA produces very different results when different subsamples were used to generate scalograms. The Rasch paradigm may address this shortcoming through the emphasis on invariant measurement, which allows people to be meaningfully compared across different sets of items and vice versa (Andrich, 2004).

Rasch/Guttman Scenario Scales

RGS scales are a relatively new phenomenon within the scale development literature and their creation is currently concentrated among a small group of researchers. The RGS methodology itself, detailed in Ludlow et al. (in press), follows a series of seven steps. How these steps manifested in the context of my dissertation instrument will be discussed extensively in Chapter 3; however a brief overview is provided here. The seven RGS scale development steps are:

1. Define the construct.
2. Determine facets and generate narrative descriptions for each facet.
3. Determine facet levels and generate descriptions to capture variation within each facet.
4. Determine the structure of the scenario instrument.
5. Develop the mapping sentences and construct the scenario items.

6. Decide on response options and survey instructions.
7. Test congruence of theory and practice.

The first three steps in this development process emphasize definition at various levels, beginning with the overall construct to be measured, followed by the construct facets, and finally the levels of each individual facet. These three steps are likely to be iterative and require extensive consultation of literature or subject experts. Focus groups may also be useful at this stage. Once these various definitions have been finalized, the process continues with explicating the scale structure, constructing scenario items, and rounding out the scale with response options and instructions. As with the definition phase, this is likely to be an iterative process requiring additional focus groups or pilot studies. Finally, when the scale is administered, its properties must be assessed with respect to the Rasch principles and fit criteria described earlier in this chapter.

There are currently five published or in-press RGS scale applications. These applications are a long and short-form scale measuring engagement in later-in-life activities (Ludlow et al., 2014; Ludlow et al., 2019), teachers' practice for equity (Chang et al., 2019), parental engagement (Antipikina & Ludlow, in press), and living a life of meaning and purpose (Ludlow et al., in press). Each of these applications highlights the construct definition process, the use of mapping sentences to construct items, and the evaluation/revision process. Each application also highlights the interpretability of scale scores that follows from application of the RGS methodology.

Section Summary

This section presented the details of the Rasch paradigm and facet theory design, both of which will be used to guide the development of my dissertation instrument. In addition to highlighting the individual strengths of each of the frameworks, this section also provided a compelling justification for bringing the two together in the RGS scale development approach. A brief overview of the RGS methodology and existing applications was also presented.

Chapter Summary

This chapter provided an overview of relevant bodies of literature informing the creation of an RGS scale measuring students' perceptions that faculty are available outside of class. First, the general importance of faculty-student interaction was established, followed by an in-depth examination of the specific benefits of engaging in out-of-class communication (OCC). This led to a discussion of existing instruments that measure constructs related to OCC, followed by an overview of their limitations. Finally, the Rasch paradigm and facet theory were discussed in detail, building a case for the combination of these methodological approaches. Chapter 3 will build upon this literature review by putting all of these theories into practice, providing a detailed discussion of the specific methods employed in my dissertation.

CHAPTER THREE: METHODOLOGY

This chapter discusses the methodology employed to address my dissertation's primary and secondary research questions:

1. Can Rasch/Guttman Scenario scales provide valid and reliable measurement of student perceptions of faculty availability outside of class?
2. What is the relationship between scores from these scales and students' participation in out-of-class communication?

I present my dissertation's overall design, with particular emphasis on the RGS scale item development process.

Research Design

My dissertation has four development phases: (a) definition of the construct(s), (b) development of items, including soliciting feedback from focus groups and a small pre-pilot, (c) a pilot study, and (d) a final administration. Refinement of the construct definition and revision of the scenario items is an interactive process interwoven between the first, second, and third phases. Data collected from the final administration suggests that the instrument may benefit from additional edits to the items; however, these are beyond the scope of my dissertation. The sections that follow discuss the general parameters of my study (e.g., participants, sampling) before describing each study phase in detail.

Population Definition

Measurement instruments do not gain value simply through existing; rather, the value of measures is governed by the information they provide about respondents. Respondents are as important to this process as the measurement instrument; thus, any instrument development project must clearly define a target population. The "postsecondary students" for whom my dissertation instrument is intended are not a homogenous group. This classification applies to many groups of students with diverse arrays of goals at different kinds of institutions. I use two characteristics to define the target population for my RGS scale: student status (i.e., undergraduate versus graduate or non-degree seeking) and institution type (residential, commuter, etc.). My instrument's intended population is undergraduate students attending residential colleges and universities. These characteristics were selected for target population definition because they have important implications for how students might engage in OCC, potentially affecting their beliefs regarding faculty members' availability for OCC. These implications are explained in the following paragraphs.

Undergraduate students are the focus of most existing research regarding OCC (see Chapter 2). Student-faculty interaction for graduate students is fundamentally different from such interactions for undergraduate students. One reason for this is undergraduate and graduate students have different kinds of relationships with faculty. Faculty often serve not only as educators, but also employers for graduate students (Lechuga, 2011), which may have consequences for OCC that are not as readily applicable to undergraduate students. This includes the topic of OCC. A key issue for graduate students and faculty is determination of authorship for publications (Welfare & Sackett, 2011); this is less likely to be the case for undergraduate students. These kinds of differences led me to believe that perceptions of availability may operate differently within graduate and undergraduate populations. Thus, I focus exclusively on undergraduates in this study. However, measuring graduate students' perceptions of faculty availability represents an area for future research in the broader field of student-faculty interaction.

Institutional type is also likely to have implications for OCC. For example, informal chance interactions between faculty and students may be much less likely to occur at commuter colleges simply because the institution is not designed such that faculty and students both spend extensive amounts of time within the confines of campus. Distance-learning also has consequences for OCC; students enrolled in online programs would never be expected to interact with faculty face-to-face, making that particular form of OCC irrelevant. These differing conceptions of OCC are why I focus exclusively on students at residential colleges for my dissertation instrument's target population. As with student status, adapting the instrument for different institutional contexts represents a potential area for future research.

Sampling Procedures

I recruited all participants for focus groups, the pilot study, and final administration in partnership with the Office of Institutional Research and Planning (IRP—formerly the Office of Institutional Research, Planning, and Assessment) at Boston College. For the pre-pilot study, I recruited a convenience sample through my personal network. Sample size was a key consideration for all administrations. Instrument development projects require that sample sizes be large enough to obtain item parameters that are stable and accurate. Crocker and Algina (1986) recommend a minimum sample size of 200 for an early-stage item analysis study; however, this recommendation is not specific to scales developed in the Rasch paradigm. While Rasch analysis procedures can still be performed on relatively small samples, inadequate sample size increases measurement error, which detracts from what can be confidently inferred about both the properties of the items and the scores of the individuals. However, Lord (1980) notes that the Rasch model is still more appropriate for small sample sizes compared to other IRT procedures because the estimates produced are still relatively accurate. This strengthens the argument for use of the Rasch model in small, exploratory studies. It is also worth noting that very large sample sizes may affect the approximate “t” statistic used to assess item fit in the Rasch model, artificially inflating these values (Smith et al., 2008).

Despite the lack of a hard rule for sample size, Rasch literature provides some general guidelines. Wright (1977) suggests that for instruments that are “well-positioned” (i.e., the item difficulties and person abilities are well-aligned) a sample size of 100 is typically adequate and that a sample size of 400 is “almost always enough” (p. 106). Linacre (1994) also recommends a sample size of at least 100 individuals in order to obtain item and person estimates that are stable within one-half of a logit. The guideline of at least 100 responses is echoed in Chen et al. (2014), who randomly selected samples of different sizes to perform Rasch analyses of a well-known and widely-utilized instrument in the healthcare field. The authors found that sample sizes less than 100 had a deleterious effect on the standard errors of item parameters. Specific to the rating scale model, Linacre (2002) provides a recommendation that at least 10 responses be observed for each rating scale category of each item.

My final instrument is comprised of two scales, each containing seven scenario-style items with five response options. Earlier drafts of the instrument (administered in the pre-pilot and pilot studies) had nine items with five response options. Taking into account the guidelines presented above, the target sample size for the pilot study and final administration was 150 students. This should allow for stable estimates of item parameters and is more than large enough to allow for 10 respondents to fall into each response category for each item. Even if Linacre's (2002) recommendation for the number of respondents per category is not attained, a sample of 100 individuals should be adequate for analysis based on the recommendations of other authors. This can be assessed by looking at the standard errors for item estimates. If the standard errors for the items are large and overlap, this indicates a lack of precision in estimation of the item parameters, as well as the possible disordering of items. Standard errors for item estimates will be evaluated for each round of data collection to monitor the adequacy of obtained sample sizes.

Pre-pilot Study

I conducted a small pre-pilot of my instrument using a convenience sample in late February 2019. This pre-pilot was conducted to provide early confirmation of the latent constructs' general structure rather than to provide precise estimates. Because of this, the target sample size was much smaller than for the pilot study and final administration. I hoped to receive at least 35 responses and 40 undergraduate students ultimately completed the instrument. This falls below the guidelines discussed above; however, the purpose of the pre-pilot was not to calibrate the items; it was simply to serve as a check on my general conceptualization of the construct(s) and clarity of the items.

Pilot Study

Potential respondents for the pilot study were randomly sampled from a list of enrolled undergraduate students at Boston College provided by IRP. Students enrolled in the Woods College of Advancing Studies (WCAS) were excluded from the sampling frame because they are not classified as traditional undergraduate day students. All other undergraduate students across all levels (freshmen, sophomores, etc.) were included. 1,500 students were invited to take the survey. Assuming a reasonable response rate of 10%, I suspected this would be sufficient to reach the target sample size of 150 respondents. Pilot data were collected in March-April 2019.

Final Administration

Recruitment for the final administration is similar to the pilot study; however, several adjustments were made in response to the pilot study's low response rate. 3,000 students were randomly sampled for recruitment from a list of enrolled Boston College undergraduate students provided by IRP. WCAS students were again excluded and students at all levels were eligible for participation. With this larger pool of potential respondents, a response rate of 5% would be sufficient to obtain the target number of respondents (150). Full administration data were collected in October-November 2019.

Rasch/Guttman Scenario Scale Development Process

My dissertation process ultimately results in two RGS scales intended for use in tandem to measure undergraduate students' perceptions of faculty availability outside of class. Together, these instruments are collectively referred to as the Out-of-Class Availability Scales (OCAS). This section covers the first major phase building the OCAS: initial development of the RGS instrument. Two tasks are discussed at length. First, I present the process of creating the scenario items. Second, I detail the various means of feedback solicitation that occurred prior to the instrument's final administration.

Creation of Scenario Items

RGS scales require a unique developmental approach that combines the Rasch measurement paradigm (Rasch, 1960/80) with Guttman's facet theory design (Guttman, 1959; Guttman, 1954). Ludlow et al. (in press) provide a series of steps that guide this process. The following sections present a detailed discussion of how each step was applied to the development of the OCAS. Although the steps are presented sequentially here, this process did not proceed linearly; many steps required several iterations. Feedback from the two sets of focus groups conducted with undergraduate students informed all steps of this process. These focus groups are described in more detail later in the chapter. RGS development step titles are reproduced from Ludlow et al. (in press).

Step One: Define the Construct

The first step in building the OCAS was defining and bounding the construct of *students' perceptions that a faculty member is available outside of class*. As shown by the diversity of OCC measures reviewed in Chapter 2, this construct can be defined in many ways. In my case, construct definition was driven by the desire to create an instrument yielding scores with clear interpretations that are informative for faculty evaluation or development. This made it essential that the construct be defined in such a way that faculty are able to adjust their behaviors based upon OCAS scores.

Keeping this end goal in mind, the first step in construct definition involved a thorough review of existing literature concerning OCC, as well as the measures that were employed in these studies. Two general dimensions of availability emerged from this review: a physical dimension and a social dimension. The most explicit articulations of these dimensions appear in the work of Wilson et al. (1974) and Bippus et al. (2003)

(both reviewed in Chapter 2). I initially considered these two dimensions together under a single umbrella construct of students' general perceptions regarding faculty availability. However, discussion with colleagues, as well as the overarching goal of creating an instrument providing scores with specific interpretations led me to separate the physical and social dimensions of availability into two distinct scales, each of which would yield its own score. I hypothesize that a single faculty member may vary in their scores across the two scales—for example, a faculty member may be perceived extremely socially engaged during OCC, and at the same time be perceived as rarely physically accessible in spaces where OCC occurs. The separation of the physical and social dimensions into different scales allows for a more accurate depiction of this reality. This separation is also a more faithful reflection of the facet theory design concept of the common range, as items on each scale will be clearly connected to the underlying construct (i.e., physical accessibility or social engagement). Arguing for the existence of a common range would have been more difficult if the physical and social dimensions were considered together in a single scale. These two scales that make up the OCAS are called the Physical Accessibility Scale (PAS) and the Social Engagement Scale (SES). These are the acronyms used in the remainder of this chapter and throughout the rest of my dissertation.

One important topic from the OCC literature was not included as part of students' perceptions of faculty availability as measured by OCAS: the purpose of OCC. This topic is clearly important in the broader landscape of OCC research; however, I chose not to include it as part of the availability construct. While it may be true that students are more likely to discuss a greater variety of topics with faculty whom they perceive as more available, discussion topics themselves are not a part of availability. Although excluded from the OCAS, the topics students discuss with faculty during OCC are still captured within my larger dissertation instrument and are used to address my secondary research question of how OCAS scores are related to actual participation in OCC.

There were several other latent constructs from the OCC literature that are ultimately separate from students' perceptions of faculty availability. These include students' satisfaction with OCC (e.g., Aylor & Oppliger, 2003; Jaasma & Koper, 1999) and students' perceptions regarding the importance of engaging with OCC (e.g., Faranda, 2015; Knapp, 2005). These constructs are also an important part of the OCC; however, they were not included in the OCAS because they are distinct from students' perceptions of faculty availability outside of class.

Step Two: Determine Facets and Generate Narrative Descriptions for Each Facet

After determining the two domains of students' perceptions of faculty availability (physical accessibility and social engagement), my next step was to determine the facets and discern how they might be combined to provide meaningful information in the context of scale items. In order to do this, I determined it was essential to attend to the different contexts within which OCC might occur. This is one limitation of existing measures discussed in Chapter 2—they often considered only a single form of OCC (for example, looking only at face-to-face OCC and excluding electronic communication).

To attend to this complexity, facetization of physical accessibility and social engagement occurred along the lines of contextual spaces: scheduled meetings, chance encounters, and email. These spaces were selected based on the frequency with which they were mentioned in existing literature and through consultation with my peers and focus groups. They were also contexts within which a single individual might reasonably be expected to display variation. For example, with respect to physical accessibility, a faculty member may be frequently present for scheduled meetings, but rarely respond to emails in a timely manner. Specific descriptions of each facet are provided in the following paragraphs.

Scheduled Meetings. This facet captures students' perceptions of faculty availability for OCC that occurs within planned times. This facet was originally identified as "office hours;" however, I determined this was not sufficient to capture the spirit of the facet. Although office hours may be the primary space in which these kinds of meetings occur; it is not the only option. Faculty may meet with students outside of their regularly scheduled office hours. These meetings may also take place via video conferencing software rather than face-to-face. Scheduled Meetings sufficiently encompasses all of these possibilities.

Chance Encounters. This facet is intended to capture what Cox (2011) refers to as "incidental contact," which "occurs when a student and a faculty member interact simply because they find themselves physically in the same place at the same time" (p. 51). Chance encounters occur if students and faculty happen to cross paths on campus grounds or in the halls of a campus building. This facet was initially named "Informal Chance Encounters"; however, "informal" was dropped due to concerns that some students might view all interaction with faculty as being formal.

Email. Because of increased email use, faculty's email response practices were considered an essential element in shaping students' perceptions that faculty are available outside of class. There are also other ways for students and faculty to interact electronically, such as social media (Sarapin & Morris, 2015). However, social media was excluded from the facet because this form of contact was deemed to fall outside of what might reasonably be expected from faculty members. Electronic communication may also occur through course platforms, such as Canvas. These platforms were excluded from the facet based upon feedback from focus group students, who indicated that email was by far the most popular form of electronic communication between students and faculty.

Step Three: Determine Facet Levels and Generate Descriptions to Capture Variation Within Each Facet

Following the determination of facet, my next step was to determine how faculty might vary meaningfully along a continuum of behaviors that would shape students' perceptions of their availability outside of class. This is an essential step for any project situated within the Rasch paradigm; the necessity of a clear, unidimensional continuum along which individuals may vary is emphasized throughout the methodological literature (e.g., Rasch, 1960/80; Ludlow et al., 2014). My initial conceptions of these continua are presented in Appendix A.

I then delineated high, medium, and low levels of physical accessibility and social engagement in a narrative format. These narratives are based upon faculty practices related to OCC in existing literature, as well as my own personal experience and discussion with current students. The narrative descriptions are presented in Appendix B. Once crafted, these narratives were distilled down to key words that captured the essence of what faculty behaviors or qualities lead to perceptions of high, medium, and low availability within each facet for the PAS and SES. These brief descriptors are presented in Appendix C. Together, these key words and rich descriptions served as the foundation for the mapping sentences that ultimately became the OCAS items.

Step Four: Determine the Structure of the Scenario Instrument

The PAS and SES each have three facets (arranged meetings, chance encounters, and email) with three levels (high, medium, and low), meaning that a total of 27 scenario items could be created for each scale if every possible combination of facet levels was used. I did not pursue this route for two main reasons. First, 27 scenario items would have been far too taxing for respondents, especially given the increased reading load inherent in the item format. Second, some of the resulting facet level combinations would be not conform to my theorization of physical accessibility and social engagement as hierarchical continua, and so would detract from the interpretability of scale scores.

To guide selection of a meaningful subset of possible facet level combinations, I followed the lead of Ludlow et al. (2014), who used a contrasting groups procedure to select combinations of facet levels for inclusion in their own RGS scales. This involves beginning with combinations that clearly represent individuals at high, low, and medium levels on the overall construct of interest. These items have all facets at a single level (high, medium, or low). Additional items that vary in level across facets are then incorporated as necessary and appropriate. This approach serves two purposes: (a) it ensures that the items are able to provide clear evidence of data fit to the Rasch model, which necessitates a clear unidimensional continuum, and (b) prevents the construction of items that are unclear or ambiguous to respondents (Ludlow et al., 2014). The facet level combinations included in the initial and final OCAS are presented in Appendix D.

As Tables D.1 and D.2 show, nine items were initially constructed for each scale. Three items within each scale depict faculty members who are constant in their levels across all three facets. Six additional items are interspersed among these facet level combinations, reflecting faculty members who differ in their physical accessibility or social engagement across the three facets. No items contain all three facet levels (high, medium, and low) for physical accessibility or social engagement, as these seem unlikely to occur in reality. Additionally, a large degree of facet level variation within the item would make it difficult to hypothesize where the item would fall along the construct continuum. It can also be seen in Tables D.1 and D.2 that each scale initially had two areas of redundancy with respect to item coverage of the construct continua. These planned redundancies were included in early versions of the scale because my theory of the physical accessibility and social engagement constructs did not suggest obvious facet level combinations for these areas of the continua. Once data were collected, items were dropped to reduce this redundancy. The schema for the final OCAS, each of which contains seven items, can be found in Tables D.3 and D.4. Discussion of how specific items were selected for removal is addressed in Chapter 4.

Step Five: Develop Mapping Sentences and Construct Scenario Items

The mapping sentence of facet theory design (Guttman, 1954; Guttman, 1959) is the foundation of the OCAS scenario items. Once the set of facet level combinations for inclusion in the OCAS had been determined, I created specific descriptions for each level to use within the mapping sentences. These descriptions were informed by the detailed narratives and key words detailed in Step Three. I used “unobtrusive” facetization when authoring the level descriptions, which means that facet levels were not repeated in an obvious manner across items in which they appeared (Borg & Shye, 1995, p. 34). I chose this approach to allow flexibility in the wording of the scenario items to make them more engaging and intuitive for respondents.

The OCAS presented an interesting choice with respect to mapping sentences and item writing. Should students be prompted to respond to scenario-style items depicting a student perceiving a specific faculty member to have a particular level of availability outside of class? Or should students be asked to compare a specific faculty member with a faculty member presented as having a particular level of availability within the scenario-style item? The first option might be considered a more “direct” measurement of students’ perceptions of availability; however, the information it would provide would not be particularly valuable from a faculty development perspective. Scores obtained from an instrument where students compare themselves to other students might provide an estimate of their perceptions of faculty availability, but would not provide faculty with actionable information regarding students’ perceptions of their availability. Thus, I decided to write items depicting faculty members who varied in their availability, not students who differed in their perceptions of faculty availability. Students’ perceptions of availability are embedded into the nature of the instrument, as all the responses they provide are a reflection of those perceptions.

The mapping sentence structures for the PAS and SES items are presented below. Following the format used in Borg and Shye (1995), the facets and their respective levels appear in parentheses. Names of the facets appear above the levels within the mapping sentence. The italicized text serves to link the facets together into a narrative description.

Mapping Sentence 1. Physical Accessibility

	Facet M: Arranged Meetings
	(high [sufficient & flexible])
Professor X has	(medium)
	(low [insufficient & inflexible])

times for arranged meetings with students. S/he is

Facet C: Chance Encounters	
(high [a constant presence])	
(medium)	*around campus.*
(low [never seen])	

	Facet E: Email
	(high [timely & consistent])
Professor X's responses to email are	(medium).
	(low [untimely & inconsistent])

Mapping Sentence 2. Social Engagement

	Facet M: Arranged Meetings
	(high [attentive & engaged])
Professor X is	(medium)
	(low [inattentive & disengaged])

during arranged meetings with students. S/he

Facet C: Chance Encounters	
(high [is happy to see])	
(medium)	*students around campus.*
(low [is annoyed to see])	

	Facet E: Email
	(high [helpful & professional])
Professor X's responses to emails are	(medium).
	(low [rude & unprofessional])

These mapping sentences served as the foundation in writing scenario items for the OCAS. These items underwent many rounds of internal revision focused upon enhancing clarity for respondents. Early item drafts tended to match the mapping sentences very closely; this correspondence was relaxed in subsequent versions. Appendix E juxtaposes the first and final drafts for each of the items.

One important consideration during this process was the order in which facets should appear within the scenario items. The mapping sentences present the facets in a fixed order; however, for combinations of facet levels this led to awkward scenario construction. An example from the PAS is presented below. An early draft of this item kept facets in the order that they appeared within the mapping sentence, which led to an awkward moving back and forth between higher and lower levels of facets across the item. In its final form, the order of facets was adjusted so that those depicting the same level appear consecutively within the item, enhancing clarity and readability.

Version 1. Facets in order of mapping sentences.

Professor Liu **holds office hours multiple days of the week and meets with students via videochat when necessary** (*M3*). She **is often on campus, but is not necessarily a visible presence to students because she spends most of her time in her office** (*C2*). Professor Liu **responds quickly to emails; students know they will usually receive a response within 24 hours** (*E3*).

Version 2. Facets at same level presented consecutively.

Professor Liu **holds office hours multiple days each week and meets with students via videochat if needed** (*M3*). She also **responds quickly to emails; students know they**

will often receive a response within 24 hours (*E3*). However, **students do not usually see her around campus** (*C2*).

Natural language emerged as an important consideration throughout the item revision process. As the goal of the OCAS is to provide scores that are easily interpretable, it was essential that the items be as clear as possible not only for respondents, but also for the faculty members who will ultimately interpret students' scores based on the scenario text.

Other important considerations in scenario construction were names and genders for the fictitious faculty members. Names changed several times over different iterations of the items. In their final form, the scenario faculty members are referred to as "Professor" accompanied by a last initial that varies across the scenarios. This construction was deemed to mitigate concerns of potential biases based on names implying different ethnic origins, as well as avoid the awkward repetition of "Professor X". For gender, the items were initially written using pronouns reflecting a balance of male and female fictitious faculty members. However, the gender-neutral, singular "they" pronoun is used in the final items to mitigate concerns regarding gender bias. These particular revisions are discussed in more detail in Chapter 4.

Step Six: Develop Response options and Survey Instructions

When students complete the OCAS, they are prompted to compare a specific faculty member they have in mind with the fictitious faculty member depicted in the scenario items. It is important that the sample represent a wide range of perceptions along the accessibility and engagement continua during instrument development. While it is impossible to guarantee this representation, the survey instructions are used to aid to this goal. Students completing the OCAS are prompted to answer the survey items keeping in mind the faculty member *with whom they most recently had class*. It is hypothesized that this instruction will lead to a more diverse set of faculty members represented in the students' responses, which stands a greater chance of achieving variation in scores. Students are also instructed to respond to the items keeping in mind their *typical interactions* with the faculty member; this instruction is provided to guard against undue influence of a single encounter that may not be representative of the students' whole experience with a particular faculty member. Finally, the use of the gender neutral, singular "they" pronoun was noted in the survey instructions in case some respondents were unfamiliar with its usage.

Students are presented with the following query after each scenario-style item: How does the faculty member you have in mind compare to Professor [X]? The response options and their respective numeric codes for the PAS and SES are presented below.

PAS Response Options.

- They are much more accessible than [Professor name]. (5)
- They are a little more accessible than [Professor name]. (4)
- They and [Professor name] are about the same. (3)
- [Professor name] is a little more accessible than them. (2)
- [Professor name] is much more accessible than them. (1)

SES Response Options.

- They are much more engaged than [Professor name]. (5)
- They are a little more engaged than [Professor name]. (4)
- They and [Professor name] are about the same. (3)

- [Professor name] is a little more engaged than them. (2)
- [Professor name] is much more engaged than them. (1)

The response options and survey instructions are likely to change once the OCAS has been fully developed and refined. I hope that eventually the OCAS will be incorporated into end-of-semester course evaluations; if this occurs, instructions will be modified to prompt students to consider the faculty member whose course they are evaluating. Depending on the platform of administration, the response options may also be adjusted to reference specific faculty member names.

Feedback Prior to Final Administration

Three forms of feedback of varying degrees of formality were solicited during the OCAS development process. Most formally, I conducted focus groups with undergraduate students at Boston College, a small pre-pilot study, and a pilot administration of the OCAS prior to the final administration. Second, I deliberately consulted with my dissertation committee members between each of these formal feedback rounds. Each committee member had the opportunity to review my finalized instrument and provide feedback before the final administration. Finally, I obtained less formal feedback on my construct definition and items from my departmental peers throughout the development process. Although they may not have the degrees that designate them as experts, these individuals have all been trained in survey development procedures and their contributions were extremely valuable.

Student Focus Groups

I conducted two rounds of focus groups to obtain qualitative feedback regarding the OCAS items. The first round of focus groups took place in January 2019 and the second in February 2019. Students were asked to consider the construct of faculty availability as a whole, as well as whether the facets of the items accurately represented the contexts in which they interacted with faculty outside of class. They were also asked to comment on the clarity and wording of the OCAS scenarios, response options, and instructions. A copy of the focus group protocol appears in Appendix F.

Students indicated that they interacted with faculty primarily during office hours and via email; however, they agreed that incorporating chance encounters as a facet captured a unique element of availability. Focus group participants were also asked to read through both the PAS and SES items and provide feedback on the item format and clarity. Students reported that the items were generally reasonable and that the process for responding to the items was clear. More specifically, participants expressed appreciation that all items "went in the same direction" so it was not necessary to extend mental effort to think about which response options accurately captured what they wished to convey.

Pre-pilot Study

A small pre-pilot for the OCAS took place in late February 2019. Forty undergraduate students participated in this pre-pilot. Pre-pilot results were used for several purposes: (a) providing general early evidence that the OCAS items conform to their hypothesized difficulty order and reflect the construct continuum, (b) identifying any potentially problematic items, and (c) providing early evidence of alignment between item difficulties and person abilities (i.e., are the distributions generally matched?). Results of this administration are presented in Chapter 4.

Pilot Study

A pilot administration of this dissertation instrument occurred in March-April 2019. The purpose of this pilot was to gather additional empirical evidence of the OCAS quality and inform any necessary revisions to items prior to the final administration in the Fall 2019 semester. Results of these analyses are also presented in Chapter 4.

Data Collection

Instrumentation

The final instrument I administered to students had three components: (a) the two 7-item OCAS scales, (b) additional items soliciting information about respondents' actual communication with faculty outside of class, and (c) additional items capturing students' reports of information about the faculty member they rated. Each of these three components is discussed in detail in the following sections and the full survey appears in Appendix G. The survey was administered using Qualtrics and the data collection plan was approved by the Boston College Institutional Review Board. The instrumentation described here pertains to the final administration. Students' reports of OCC participation and faculty characteristics were not a part of the pre-pilot and students' reports of faculty characteristics were not included in the pilot.

The OCAS

The backbone of this survey is the two 7-item OCAS scales: the Physical Accessibility Scale (PAS) and the Social Engagement Scale (SES). The PAS and SES each yield a score that represents a student's perception of a particular faculty member's availability. A high score indicates that a student perceived the faculty member they rated to be highly physically accessible or socially engaged outside of class. A low score indicates that a student perceived the faculty member they rated to be inaccessible or disengaged outside of class. It is expected that the items deliberately authored to capture high levels across the facets within each domain will be more difficult for students to answer positively (i.e., to select a response option indicating that the faculty member they rated is *more* accessible or engaged than what is presented in the scenario) than items authored to reflect low facet levels within each domain.

The OCAS measure students' *perceptions* of faculty availability. These perceptions may differ from a faculty member's perceptions of their own availability, or even from their actual availability. The scenario items should serve to mitigate discrepancies between students' perceptions and actual availability because they provide detailed descriptions against which students compare the faculty member they are rating. The representation of specific behaviors within the items allow students to make more concrete comparisons, as opposed to those in more typical Likert items (e.g., My professor is physically accessible outside of class). Discrepancies between students' perceptions and faculty members' own perceptions may also occur. While the scenario items cannot prevent this, they can provide a more concrete starting point for investigating these discrepancies. For example, rather than a faculty member simply being miffed at being rated low on "availability" from a traditional course evaluation, she could consult the Rasch variable map accompanying an OCAS score and ascertain a clearer picture of how she is being perceived by students.

The purpose of my dissertation is to develop the OCAS instrument, not to provide interpretable feedback for specific faculty members, so it is not an issue that all students are considering different faculty members when completing these items. However, this may not be the case for future research using the OCAS. Discussion of how OCAS scores might be presented to and used by individual faculty members or departments is presented in Chapter 5.

Students' Actual Participation in Out-of-Class Communication

I included items related to students' actual participation in OCC with the faculty member they rated in order to address my dissertation's secondary research question: What is the relationship between scores from the OCAS and students' engagement in OCC?

Respondents were asked to estimate the number of interactions they had outside of class with the faculty member they rated in different contexts. These frequencies were compared with scores from the PAS and SES to test whether or not student perceptions of faculty availability outside of class are positively associated with OCC participation. While not necessarily informative for refinement of the OCAS items, this information supports the assertion that students' perceptions of faculty availability outside of class are key in determining whether or not students engage in OCC. Respondents were also asked to indicate the purposes for which they engaged in OCC with the focal faculty member. These responses are used to investigate whether or not higher scores on the PAS or SES are associated with discussing particular topics with the faculty member during OCC. Finally, students completed a single Likert-type item indicating their overall satisfaction concerning OCC with the faculty member they had in mind while completing the OCAS. Responses to this item are used to examine the relationship between students perceptions of availability and OCC satisfaction.

Faculty Characteristics

The final component of my survey instrument asked students to provide contextual information about the faculty member whom they rated. Specifically, students were asked to indicate their perceptions of the faculty member's gender, race, and age. They were also asked to report the department within which the faculty member works. This information will not be used within my dissertation study itself; however, it will be used in future work to examine how faculty characteristics may influence students' perceptions of physical accessibility or social engagement.

Data Analysis

Data collected from the OCAS pre-pilot, pilot, and full administration are all analyzed using the Rasch rating scale model (Andrich, 1978). These analyses produce several important pieces of information that are essential for evaluating if students' perceptions of faculty availability outside of class can be adequately captured through the use of Rasch/Guttman scenario scales. The rating scale model is presented in the following equation:

$$\pi_{nix} = \frac{e^{\sum_{j=0}^{x_{ni}}[\beta_n - (\delta_i + \tau_j)]}}{\sum_{k=0}^{m} e^{\sum_{j=0}^{x_{ni}}[\beta_n - (\delta_i + \tau_j)]}} \quad (7)$$

Where:

- π_{nix} is the probability of person *n* responding to item *i* in category *x*.
- β_n is the "ability" estimate for person *n* (Note that the definition of "ability" varies depending upon the context).
- δ_i is the estimate of difficulty for item *i*.

- τ is the threshold difficulty for the *k*th step from one response category to the next.

The Rasch rating scale analysis provides estimates for both respondents and items along the physical accessibility and social engagement continua. Estimates for respondents will indicate how accessible or engaged students' perceived their particular faculty member to be; a high estimate would show that a student perceived a faculty member to be very accessible or engaged, while a low estimate would show that a student perceived a faculty member to be inaccessible and disengaged. For items, a high estimate would indicate that the item reflects a high level of accessibility/engagement and is more "difficult" for students to rate faculty members highly on (i.e., to select a response option indicating that their faculty member is *more* accessible or engaged than what is depicted in the scenario). Low item estimates show that an item reflects relatively low levels of accessibility/engagement and is less "difficult" for students to rate faculty members highly on.

Because these items were purposefully authored to reflect varying levels of accessibility or engagement, they should spread along the continuum if data collected fit the Rasch model. More specifically, the items written to depict very high perceived levels of accessibility or engagement should be the items where it is most difficult for students to perceive their focal faculty member as *more accessible* or *more engaged*, while those depicting very low levels of accessibility or engagement should be the items where it is least difficult for students to do so.

The degree to which the OCAS items fit the hypothesized continuum is reflected in the variable maps produced through the Rasch analyses. As discussed in Chapter 2, items that are intended to be most difficult should appear at the top of the map, while items that are easiest should appear at the bottom. Additionally, while there may be clusters of items that are designed to reflect similar levels of accessibility/engagement, there should not be large gaps between clusters of items. Large gaps would indicate areas of the physical accessibility or social engagement constructs that are not adequately captured within the scale.

I also examined additional Rasch analysis output to evaluate the OCAS. The OCAS category characteristic curves (CCCs) are evaluated to ensure that all response options are utilized across the items. Fit statistics are also assessed to provide additional evidence of data-model fit. All analyses are completed in the Rasch analysis software WINSTEPS (Linacre, 2018).

As discussed at length in Chapter 2, a distinguishing feature of the Rasch paradigm is its emphasis on whether or not collected data fit the model (Andrich, 2004). All of the analyses described in the preceding paragraphs are means of assessing this requirement. Each set of analyses throughout my dissertation (pre-pilot, pilot, and full administration) is conducted with the ultimate goal of providing evidence that data gathered through the OCAS fit the Rasch model. As noted by Andrich (2004), lack of congruence between that data and model in the Rasch paradigm is indicative of potential problems regarding the conceptualization of the construct, or specific items/persons. Adequate data/model fit provides evidence for the appropriateness of the OCAS' conceptualization of the physical accessibility and social engagement constructs, while poor data/model fit indicates the need for revisions of either the overall construct definition or specific problematic items.

Chapter Summary

This chapter has outlined the specific procedures I employ in my dissertation. I have presented the target population for the OCAS instrument, as well as a justification for its selection. Information regarding sampling for each round of data collection is also provided. Perhaps most importantly, I have provided detailed documentation of the OCAS development process and the various types of feedback I solicited throughout that process. Given the relatively small number of existing RGS scales, this documentation in and of itself is a contribution to the scale development literature. Finally, my full survey instrument and analysis procedures were also described.

CHAPTER FOUR: RESULTS

This chapter presents the results from my three rounds of data collection to address my dissertation's primary and secondary research questions:

1. Can Rasch/Guttman Scenario scales provide valid and reliable measurement of student perceptions of faculty availability outside of class?
2. What is the relationship between scores from these scales and students' participation in out-of-class communication?

Development of the Out-of-Class Availability Scales (OCAS) was a lengthy process requiring multiple administrations of scale items. The same Rasch analysis procedures were performed for each round of data collection, making the results presented here somewhat repetitive. Improvements in the scales' properties with each successive administration are emphasized. Some material related to the presentation of Rasch results (e.g., extensive discussion of how a variable map is formatted) is presented only once to reduce redundancy.

Pre-Pilot Study

Overview of Responses

The pre-pilot study consisted of a convenience sample. As noted in Chapter 3, the purpose of this pre-pilot administration was to obtain empirical support for my theorization of the construct continua. Small convenience samples are suitable for this purpose (Lord, 1980). Respondents were recruited by a faculty member in my personal network at a small Midwestern university. The pre-pilot was administered using Qualtrics and the faculty member made the instrument available to approximately 80 students through an anonymous link. Data were collected over a two-week period in February 2019.

Missing Data

Qualtrics captured 54 pre-pilot responses for the pre-pilot. Of these 54 records, 11 (25.6%) were completely blank (i.e., no item responses were recorded). These 11 records were deleted at the outset of pre-pilot analyses, yielding a starting sample size of 43. Of the 43 records containing data, 41 (95.3%) were complete (i.e., all nine items for both scales had recorded responses). For the three cases that were not complete, two (4.6% of total) records contained complete data for the Physical Accessibility Scale (PAS), but no data for the Social Engagement Scale (SES). The final record contained missing data for one item on the PAS and one item on the SES. A summary of missing data from the pre-pilot administration is presented in Table 4.1

Table 4.1

Missing Data Summary (Pre-pilot)

Physical Accessibility Scale			
Number of Missing Variables	Frequency (Number of Cases)	Percent	Cumulative Percent
1	1	2.3	2.3
0	42	97.7	100.0
Social Engagement Scale			
Number of Missing Variables	Frequency (Number of Cases)	Percent	Cumulative Percent
9	2	4.6	4.6
1	1	2.3	6.9
0	40	93.0	100.0

This is a small amount of missing data. The two respondents who completed only the PAS items likely quit the Qualtrics survey after completing the first screen, as the two scales were presented on two different pages. These individuals were retained for the

Rasch analysis of the PAS and excluded from the SES. This leaves the two scales with slightly different sample sizes for the pre-pilot analyses: 43 individuals for the PAS and 41 individuals for the SES. Four respondents selected 5 (the highest category) for all 9 items on the SES, so they are automatically excluded from the Rasch analyses but are included in the descriptive statistics.

A single respondent skipped one item within each scale. The missing items are not suggestive of a pattern. For the PAS, the respondent did not complete the item designed to reflect a high level across all three facets. For the SES, the respondent did not complete an item that was designed to reflect a medium level for two facets and a high level of the third facet. As items were randomized during administration, it is not possible to tell when the respondent encountered the skipped items. This individual was retained for the pre-pilot Rasch analyses because the Rasch estimation procedure is capable of addressing missing data.

Descriptive Statistics

Respondents

As noted above, respondents for the pre-pilot were a convenience sample recruited from a Midwestern university who completed the survey through an anonymous link. Unlike students from the pilot study and final administration recruited from Boston College, I did not have access to demographic data for these individuals, so this information is unknown for pre-pilot participants.

Scenario Items

Table 4.2 presents the means and standard deviations for each PAS and SES item. Items appear in their hypothesized order from most difficult (i.e., the hardest scenario for respondents to rate their faculty member as *more* accessible/engaged than the description) to least difficult. “Difficulty” is used throughout this chapter as it is in the Rasch literature to refer to an item’s position along the construct continuum. Within Table 4.2, “H” indicates a high facet level, “M” indicates a medium facet level, and “L” indicates a low facet level. Although scenario text may have varied the order in which facets appeared, for reporting purposes the first level always represents the arranged meetings facet, the second level always represents the chance encounters facet, and the third level always represents the email facet. For example, the item labeled “HHM” was written to reflect a high level for the arranged meetings and chance encounters facets and a medium level for the email facet. The number in parentheses indicates the overall “score” of the item as calculated by summing the individual facet levels. A high facet level is assigned a 3, a medium facet level is assigned a 2, and a low facet level is assigned a 1. Items are numbered within the table for reference purposes only; items were randomized when presented to respondents.

Item responses were coded according to the following scheme:

- 5 = much more accessible/engaged than X
- 4 = a little more accessible/engaged than X
- 3 = about the same as X
- 2 = X is a little more accessible/engaged
- 1 = X is much more accessible/engaged

This means that items with high means were the least difficult for respondents to select *much more accessible/engaged than X*. In contrast, the items with the lowest means are those for which it was most difficult for respondents to select *much more* *accessible/engaged than X*. For the PAS, the item with the highest mean is Item 9 (LLL) and for the SES, the item with the highest mean is Item 5 (MMM). This means that these two items were the least difficult for respondents to select the highest level option of *much more accessible/engaged than X*.

Examination of the item means in conjunction with their designated facet levels provides a preview of what will later be tested by the Rasch analysis. Generally, items with higher facet levels have lower means, meaning that it is more difficult for respondents to select the higher response categories. The items constructed to reflect low facet levels tend to have higher means, meaning that it is easier for respondents to select the higher response categories. This provides some preliminary evidence that the hypothesized construct continua for the PAS and SES function as intended. However, there are a few items that deviate from this intended progression—most notably Item 5 (MMM) in the SES. Item 5 was designed to capture a medium level across all facets, but has the highest mean. This suggests that the item was easier than intended and provides some early evidence that revisions are required.

Table 4.2

Descriptive Statistics for Scenario Items (Pre-pilot)

Physical Accessibility Scale		
	Mean	Standard Deviation
Item 1: HHH(9)	3.29	0.94
Item 2: HHM (8)	3.98	1.06
Item 3: HMH (8)	3.28	0.93
Item 4: MMH (7)	4.02	0.91
Item 5: MMM (6)	4.35	0.92
Item 6: MLM (5)	4.21	1.10
Item 7: MML (5)	4.37	0.87
Item 8: LLM (4)	4.30	0.99
Item 9: LLL (3)	4.53	0.74
Social Engagement Scale		
	Mean	Standard Deviation
Item 1: HHH(9)	3.22	1.24
Item 2: HHM (8)	3.51	0.98
Item 3: HMH (8)	3.78	0.88
Item 4: MMH (7)	4.15	0.77
Item 5: MMM (6)	4.46	0.81
Item 6: MLM (5)	4.17	0.95
Item 7: MML (5)	4.00	1.16
Item 8: LLM (4)	4.22	0.96
Item 9: LLL (3)	4.32	0.93

Rasch Analyses

Following examination of the descriptive statistics, I performed a Rasch rating scale analysis (Andrich, 1978) on the pre-pilot data for each scale. This provides a more nuanced assessment of congruence between my hypothesized theory of the perceived physical accessibility and social engagement constructs and the actual item responses provided by students. The descriptive statistics presented in the previous section provide some preliminary evidence regarding the construct continua, which is verified and enhanced by the Rasch analysis. Items will appear in the exact same difficulty order (i.e., the same order of means from highest to lowest); however, the Rasch results are accompanied by a visual representation and series of fit statistics that allow for a more detailed investigation of how the empirical data correspond (or do not correspond) to the hypothesized construct continua.

At the pre-pilot stage, my focus is upon demonstrating proof-of-concept for the continua of the two dimensions. Thus, the primary piece of Rasch output with which I am concerned is the variable map. The items designed to represent the highest perceptions of accessibility or engagement (the hardest items) should appear at the top of the variable map, while those representing the lowest (the easier items) should appear at the bottom of the variable map. This is somewhat counterintuitive, given that because of the response option coding, the most difficult items had the lowest means and the easiest items had the highest means. The basic variable map is also complemented by a variation presenting the Andrich thresholds. Both of these are explained in detail for each scale below. In addition to the variable maps, I also examined fit statistics, category characteristic curves, and the reliability and separation indices for each scale. Finally, I performed a principal components analysis on the scale residuals to obtain evidence of unidimensionality.

Variable Maps

Physical Accessibility Scale. Figure 4.1 presents the pre-pilot PAS variable map. The line running up the center of the map represents the hierarchical continuum of students' perceptions of a faculty member's physical accessibility outside of class. Higher representations and perceptions of physical accessibility are represented at the top of the map and lower perceptions of physical accessibility are represented at the bottom. Specific features of the variable map are explained in the following paragraphs. This information is not repeated for subsequent variable maps.

The “Measure” column on the left side of the map provides estimates of item difficulty and person status in the Rasch logit unit (see Chapter 2 for an explanation of this unit). Person indicators are on the left side of this line represented by three-digit numbers. The labels “<more>” and “<less>” on the person side indicate the relative positions for having “more” or “less” of the construct of interest. In this case, individuals near the top of the variable map perceive the faculty member that they rated to be “more” accessible, while those near the bottom of the variable map perceive the faculty member they rated to be “less” accessible.

Item indicators are to the right of the central line represented by their facet levels in the same manner as the previous section. The overall item level obtained by summing the individual facet levels appears in parentheses. The labels “<rare>” and “<freq>” at the top and bottom of the item side of the variable indicate rarer or more frequent “success” on items. This means that selecting a higher response option (i.e., indicating that the faculty member one had in mind is more accessible than what is depicted in the scenario) is “rarer” for the items near the top of the variable map than for those near the bottom. This is simply another way of saying that items near the top of the map are more difficult to rate highly than those near the bottom.

Means and standard deviations (in the logit unit) are noted on the variable map for both persons and items. Means for persons and items are indicated by the “M”s appearing on each side of the central line. The position of the first standard deviation is represented by “S” and the position of the second standard deviation is represented by “T”.

Figure 4.1

Physical Accessibility Scale Variable Map (Pre-pilot)

```
MEASURE                   PERSON - MAP - ITEM
                              <more>|<rare>
    5                     114  139  +
                                    |
                                    |
                                    |
                               103  |
                                    |
    4                              T+
                                    |
                               138  |
                                    |
                                    |
                     115  129  130  |
    3                               +
      124  128  133  134  135  142 S|
                                    |
                          118  140  |
                                    |
                     116  122  143  |
    2                     100  120  +
                                    |
                108  112  117  125 M|T
                     110  111  136  |  HHH (9)    HMH (8)
                     107  119  127  |
                               131  |
    1                               +
                               104  |S
                     102  126  137  |
                          105  113 S|
                          109  141  |
                               101  |  HHM (8)   MMH (7)
    0                               +M
                               132  |
                               123  |  MLM (5)
                                    |  LLM (4)
                                   T|  MML (5)   MMM (6)
                               121  |S
   -1                               +
                                    |  LLL (3)
                                    |
                                    |
                                    |T
                                    |
   -2                               +
                              <less>|<freq>
```

As can be seen in Figure 4.1, the items generally follow the hypothesized progression from low levels of perceived physical accessibility to high levels of perceived physical accessibility. Item HHH is at the top of the variable map on the item side, although it has nearly the same average difficulty as Item HMH. The relative position of these two items indicates that they are those for which it is most difficult for respondents to select *much more accessible than X*. Item LLL is at the bottom of the variable map, meaning that it is the item for which it is easiest for respondents to select *much more accessible than X*.

The item distribution is murkier in the middle of the scale. There are several items that do not fall in their hypothesized ordering (i.e., levels 4 through 6). Another noteworthy issue is the close grouping of the items. Rasch measurement principles mandate that items should be spread equally along a hierarchical continuum in order to adequately capture a construct. The variable map results here suggest that items might be revised to be more different from one another and increase their location spread. This would allow for increased confidence in the difficulty order of the items and provide further evidence for the conceptualization of the physical accessibility continuum. Given that the purpose of this administration was simply to obtain proof-of-concept for the construct, these results are encouraging even though this sample size is quite small.

Another takeaway from this variable map is that, on average, the item estimates are below the person estimates. This is illustrated by the relative positions of the "M" markers on either side of the vertical line. The "M" to the left of the line represents the average person estimate (1.88), while the "M" to the right of the line represents the average item estimate (set to 0.0). This discrepancy between person and item estimates can also be evaluated visually; many individuals appear in the upper half of the map, while most items are concentrated in the lower half. This lack of congruence between people and items means that with the pre-pilot items, it is not possible to describe what a very high perception of physical accessibility "looks like" because there are no items in that upper area of the scale.

The pre-pilot person separation for the PAS was 1.81 with a reliability of .77. This suggests that the items may not be sensitive enough to differentiate between students with high and low perceptions of social engagement. The SES pre-pilot item separation was 3.37 with a reliability of .92. This indicates that some degree of confidence is appropriate regarding the differentiation among items reflecting high and low levels of physical accessibility; however, this index will not be useful for comparison across administrations because it is a function of the number of respondents.

Figure 4.2

Physical Accessibility Scale Variable Map with Andrich Thresholds (Pre-pilot)

```
MEASURE  PERSON - MAP - ITEM - Andrich thresholds (modal categories if ordered)
         <more>|<rare> 2        3        4        5
    5      XX  +
               |
               |
               |
            X  |
               |
               |
    4         T+
               |
            X  |
               |
               |
               |                                              HHH     .5
                                                              HMH     .5
          XXX  |
    3          +
              S|
       XXXXXX  |
               |
           XX  |
               |                                HHH     .4
                                                HMH     .4
          XXX  |                                              HHM     .5
    2      XX  +                                              MMH     .5
               |
         XXXX M|T
          XXX  |                                              MLM     .5
               |                                              LLM     .5
          XXX  |                                              MMM     .5
                                                              MML     .5
            X  |                                HHM     .4
    1          +                                MMH     .4
            X  |S
               |                                              LLL.5
         XXXX S|                                MLM     .4
            X  |                  HHH     .3    LLM     .4
                                  HMH     .3
           XX  |                                MMM     .4
            X  |                                MML     .4
    0          +M
            X  |
               |    HHH     .2                  LLL     .4
                    HMH     .2
            X  |
              T|
               |
            X  |S                 HHM     .3
   -1          +                  MMH     .3
               |
               |                  MLM     .3
               |
               |    HHM     .2    LLM     .3
                    MMH     .2
               |T                 MMM     .3
                                  MML     .3
               |
   -2          +    MLM     .2
               |    LLM     .2    LLL     .3
               |    MMM     .2
               |    MML     .2
               |
               |
               |    LLL     .2
   -3          +
         <less>|<freq> 1        2        3        4
```

Figure 4.2 shows a variable map variation that includes the Andrich thresholds. In the rating scale model, Andrich thresholds represent the point at which an individual has a 50% probability of selecting the next higher response category for the item (see Chapter 2 for details). The general schematic of this variable map is the same as Figure 4.1; the differences are in how items and people are represented. In this map, each item appears four times: once for each response category transition. The item labels followed by ".2" are the Andrich thresholds for moving between response categories one and two (*X is much more accessible* and *X is a little more accessible*), labels followed by ".3" are the Andrich thresholds for moving between response categories two and three (*X is a little more accessible* and *about the same as X*), and so on.

Instead of a three-digit number, each person in this variable map is represented by an X. The placement of individuals relative to the items indicates which response categories they were likely to select. For example, an individual appearing directly across the continuum line from HMH.4 is expected to select either response category 3 (*about the same as X)* or response category 4 (*a little more accessible than X*).

It is clear from Figure 4.2 that respondents were frequently expected to select the upper two response categories for all items. Only for the most difficult items were any individuals expected to fall below the top three response categories. This provides another piece of evidence that the PAS would benefit from more difficult items and increased spread within the scale.

Social Engagement Scale. Figure 4.3 presents the SES variable map. The map has the same format as that of the PAS. The line running up the center of the map represents the hierarchical continuum of students' perceptions of a faculty member's social engagement outside of class. Higher representations and perceptions of social engagement are represented at the top of the map and lower perceptions of social engagement are represented at the bottom.

Figure 4.3

Social Engagement Scale Variable Map (Pre-pilot)

```
MEASURE              PERSON - MAP - ITEM
                         <more>|<rare>
    5      103  114  124  139  +
                               |
                               |
                          129  |
                               |
                               |
    4                          +
                               |
                              T|
                               |
                               |
                     120  138  |
    3                          +
                          135  |
                               |
                     116  134 S|
                               |
      115  118  128  133  140  |
    2      111  122  125  130  +
                               |
                          136  |
                104  107  109 M|T
           100  117  119  127  |  HHH (9)
                          121  |
    1                     132  +
                     108  110  |  HHM (8)
                               |S
                          126  |  HMH (8)
                     112  123 S|
      101  102  105  131  141  |
    0                          +M MML (5)
                          137  |  MMH (7)
                               |  MLM (5)
                               |  LLM (4)
                               |S LLL (3)
                          113 T|
   -1                          +  MMM (6)
                         <less>|<freq>
```

While the items generally follow the hypothesized progression from low to high levels of perceived social engagement, there are some problematic areas within the scale. Notably, Item LLL is not at the bottom of the variable map, despite theoretically being the easiest item. Instead, Item MMM, which was hypothesized to be in the middle of the scale, is the easiest. Further investigation of Item MMM is clearly warranted.

The SES items are also clustered closely together along the scale. This violates Rasch measurement principles, which propose that items should be spread equally along a hierarchical continuum. This suggests that items might be revised to be more different from one another and increase spread. This would allow for increased confidence in the difficulty order of the items and provide further evidence for the conceptualization of the social engagement continuum.

As with the PAS, on average, the item estimates are below the person estimates. This is illustrated by the relative positions of the “M” markers on either side of the vertical line. The average person estimate is 1.86, while the average item estimate is set to 0.0. This discrepancy is also visually apparent in the placement of individuals and items on the variable map. Four individuals had “perfect” scores on the SES, meaning that they selected 5 (*much more engaged than X*) for all nine items. This lack of congruence between people and items means that with the pre-pilot items, it is not possible to describe what a very high perception of social engagement “looks like” because there are no items in that area of the scale.

The pre-pilot person separation for the SES was 1.93 with a reliability of .79. This suggests that the items may not be sensitive enough to differentiate between students with high and low perceptions of social engagement. The SES pre-pilot item separation was 2.88 with a reliability of .89. This suggests that it would be beneficial to administer the instrument to a larger sample of respondents to increase confidence in differentiation of the construct continuum through the SES items, although item separations from future administrations will not be comparable to these numbers due to differences in sample size.

Figure 4.4

Social Engagement Scale Variable Map with Andrich Thresholds (Pre-pilot)

```
MEASURE  PERSON - MAP - ITEM - Andrich thresholds (modal categories if ordered)
         <more>|<rare> 2       3       4       5
    5    XXXX  +
               |
               |
            X  |
               |
               |
    4          +
               |
              T|
               |                                                HHH     .5
               |
           XX  |
    3          +                                                HHM     .5
            X  |
               |                                                HMH     .5
           XX S|
               |
        XXXXX  |                                                MML     .5
    2    XXXX  +                                                MMH     .5
               |                                                MLM     .5
            X  |                                  HHH     .4    LLM     .5
          XXX M|T                                               LLL     .5
         XXXX  |
            X  |                   HHH      .3    HHM     .4    MMM     .5
    1       X  +
           XX  |
               |S                  HHM      .3    HMH     .4
            X  |
           XX S|                                  MML     .4
        XXXXX  |                   HMH      .3
    0          +M                                 MMH     .4
            X  |                   MML      .3    MLM     .4
                                                  LLM     .4
               |                   MMH      .3
               |                   MLM      .3    LLL     .4
               |S                  LLM      .3
            X T|    HHH      .2    LLL      .3    MMM     .4
   -1          +
               |
               |    HHM      .2    MMM     .3
               |T
               |
               |    HMH     .2
   -2          +
               |    MML     .2
               |
               |    MMH     .2
                    MLM     .2
               |    LLM     .2
               |    LLL     .2
   -3          +
               |
               |    MMM     .2
               |
               |
               |
   -4          +
         <less>|<freq> 1       2       3       4
```

Figure 4.4 highlights that respondents were frequently expected to select the upper two response categories for all items. Almost all individuals were expected to select one of the upper response categories for most items. This provides another piece of evidence that the SES would benefit from revisions to better differentiate the items and increase overall scale difficulty.

Fit Statistics

As noted in the previous chapters, a key consideration in any Rasch analysis is the extent to which empirical data "fit the model". Rasch model fit statistics complement the variable map and can be used to flag potentially problematic individuals or items. Fit statistics for items are the primary concern during instrument development because the goal at this stage is not to provide precise estimates for individuals, so person-fit statistics are not presented here.

Two fit statistics are discussed here. The information-weighted fit statistic (MNSQ-INFIT) is a measure of consistent inconsistency, meaning that overall, students did not respond to an item in a consistent fashion. The unweighted fit statistics (MNSQ-OUTFIT) indicate individual surprising responses to items by particular students. A criterion of 1.3 is used for these statistics. Each fit statistic is accompanied by a rough "t" statistic (ZTSD), for which a criterion of +/-2 is used. Additional details regarding these indices are presented in Chapter 2.

Physical Accessibility Scale. Table 4.3 presents the fit statistics for the pre-pilot PAS data. The items are listed in difficulty order, as indicated in the logit estimate column. A higher logit estimate indicates an item for which it is more difficult for respondents to select the upper-level categories (i.e., to indicate that the faculty member they have in mind is at least as accessible as the fictitious faculty member in the item).

Table 4.3

Physical Accessibility Scale Fit Statistics (Pre-pilot)

Item	Logit Estimate (S.E.)	Information-Weighted Fit Statistic		Unweighted Fit Statistic	
		MNSQ	ZSTD	MNSQ	ZSTD
Item 1: HHH (9)	1.44 (.20)	1.64	2.61	1.63	2.56
Item 3: HMH (8)	1.42 (.19)	.83	-.79	.81	-.92
Item 2: HHM (8)	.20 (.21)	1.08	.43	1.02	.17
Item 4: MMH (7)	.11 (.22)	.66	-1.65	.81	-.78
Item 6: MLM (5)	-.28 (.23)	1.55	2.12	1.26	.99
Item 8: LLM (4)	-.49 (.24)	.83	-.68	.69	-1.15
Item 5: MMM (6)	-.61 (.24)	1.07	.36	.83	-.49
Item 7: MML (5)	-.67 (.24)	1.04	.25	.78	-.68
Item 9: LLL (3)	-1.12 (.27)	.83	-.60	.63	-1.05

Seven of the nine scale items have acceptable fit statistics, which provides evidence that the items are generally a good fit to the Rasch model. Two highlighted items in Table 4.3 had both INFIT and OUTFIT statistics greater than the criterion of 1.3. Item 1 (HHH) is the most difficult item on the scale, suggesting that the source of misfit may be low-scoring individuals who gave unexpectedly high responses. As Item 6 (MLM) is in the middle of the scale, it is more difficult to hypothesize the cause of its misfit without examining observed response patterns.

I examined the person-response table to further evaluate the two items exhibiting misfit. The person-response tables for the pre-pilot can be found in Appendix G. Examination of this table revealed that for Item 1, constructed to represent a high level across all facets, several lower-scoring respondents unexpectedly selected the higher response categories. The opposite was true for Item 6; unexpected responses occurred when individuals chose lower-than-expected response options. Inspection of the person-response tables did not lead to any specific hypotheses regarding the reason for this misfit.

Social Engagement Scale. Table 4.4 presents the fit statistics for the pre-pilot SES data. Again, the items are listed in difficulty order and a higher logit estimate indicates an item for which it is more difficult for respondents to select the upper-level categories (i.e., to indicate that the faculty member they have in mind is at least as engaged as the fictitious faculty member in the item).

Table 4.4

Social Engagement Scale Fit Statistics (Pre-pilot)

Item	Logit Estimate (S.E.)	Information-Weighted Fit Statistic		Unweighted Fit Statistic	
		MNSQ	ZSTD	MNSQ	ZSTD
Item 1: HHH (9)	1.37 (.20)	1.31	1.35	1.28	1.24
Item 2: HHM (8)	.89 (.20)	1.03	.21	1.04	.27
Item 3: HMH (8)	.43 (.21)	.88	-.47	.82	-.76
Item 7: MML (5)	.02 (.22)	1.19	.85	1.05	.28
Item 4: MMH (7)	-.20 (.23)	.66	-1.54	.64	-1.53
Item 6: MLM (5)	-.33 (.23)	1.09	.45	.88	-.41
Item 8: LLM (4)	-.44 (.24)	.98	.00	.90	-.32
Item 9: LLL (3)	-.67 (.25)	1.32	1.21	1.20	.76
Item 5: MMM (6)	-1.07 (.27)	1.06	.30	.88	-.29

Seven of the nine scale items have acceptable fit statistics, indicating that the items generally conform to the Rasch model. Two highlighted items in Table 4.3 had INFIT statistics greater than the criterion of 1.3. None of the items had OUTFIT statistics greater than the 1.3 criterion, although one item came close (Item 1: 1.28). This suggests that the source of misfit for these items is generally not surprising individual responses;

rather, the issue may be related to inconsistency of responses across students in the sample.

I examined the person-response table (see Appendix H) to further evaluate the three items exhibiting misfit. As expected, there were not clear patterns revealing the source of the misfit. Individuals did not appear to score unexpectedly high or unexpectedly low in a consistent manner for any of the three items. Rather, the high INFIT statistics are a result of general inconsistency in responses. Inspection of the person-response tables did not reveal a clear cause of misfit.

Category Characteristic Curves

Figures 4.5 and 4.6 present the category characteristic curves (CCCs) for each of the OCAS. The CCCs show the probability that an individual will select a specific response category based upon the difference between the person and item estimates in the Rasch logit unit (see Chapter 2 for details). These differences are shown on the *x*-axis and the response option selection probabilities are shown on the *y*-axis.

Physical Accessibility Scale. Figure 4.5 presents the PAS CCC. The figure shows that all five response categories have the highest probability of being selected at different points along the construct continuum. While response categories 2 and 4 (*X is a little more accessible* and *a little more accessible than X*, respectively) are not chosen as frequently as the other three categories, there are still areas of difference between person and item estimates where they are most likely to be chosen. This provides some early evidence for the appropriateness of five response categories for the PAS because it shows that all five categories are being used by respondents.

The intersections of each response category curve represent the Andrich thresholds, which were also illustrated in the variable map presented in Figure 4.2. These thresholds indicate the logit estimate where an individual has a .5 probability of selecting one response category over another. For example, the Andrich threshold for response category 2 (*X is a little more accessible*) is the logit estimate where individuals have a .5 probability of selecting category 2 or category 1 (*X is much more accessible*). These thresholds, associated INFIT and OUTFIT statistics, along with the average logit estimates of individuals making using of each category, are presented in Table 4.5. Unsurprisingly, the threshold is lowest for the transition between categories 1 and 2, and is highest for the transition between categories 4 and 5. The average estimates are also in intuitive order; response category 1 has the lowest average estimate and response category 5 has the highest. These numerical indicators corroborate what is shown visually in the category characteristic curve. However, both fit statistics for category 1 (*X is much more accessible*) are greater than 1.3, indicating both general inconsistency and particularly surprising responses. The relatively frequency of use for this response category likely contributes to this misfit.

Figure 4.5

Physical Accessibility Scale Category Characteristic Curve (Pre-pilot)

```
  CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+------+------+------+------+------+------+------+------+-
R  1.0 +                                                         +
O      |                                                         |
B      |11                                                      5|
A      |  111                                                555 |
B   .8 +     11                                            55    +
I      |       11                                        55      |
L      |         1                                      5        |
I      |          1                                   55         |
T   .6 +           11                                5           +
Y      |             1             33               5            |
    .5 +              1         333  333           5             +
O      |               1       3        33  44444 5              |
F   .4 +               2*222233           *4     *44             +
       |            222  1  3222        44 33   5   44           |
R      |          22      13    2     44     355      44         |
E      |        22       331     22  4       53         44       |
S   .2 +      22        3   11     **       5  33         44     +
P      |   222        33      1  44  22   55     33         444  |
O      |222        333         **      2*5         33          44|
N      |       3333       44444  111*555 22222       33333       |
S   .0 +******************5555555555 111111111*******************+
E      -+------+------+------+------+------+------+------+------+-
       -4     -3     -2     -1      0      1      2      3      4
        PERSON [MINUS] ITEM MEASURE
```

Table 4.5

Physical Accessibility Scale Andrich Thresholds and Average Estimates (Pre-pilot)

Response Category	Andrich Threshold	Response Frequency	INFIT	OUTFIT	Average Estimates
1 (X is much more accessible)	N/A	8	1.92	2.06	-.41
2 (X is a little more accessible)	-1.72	23	1.15	1.10	-.01
3 (about the same as X)	-1.05	81	.93	.75	.62
4 (a little more accessible than X)	.88	108	1.21	.83	1.41
5 (much more accessible than X)	1.88	166	.86	.89	2.88

Social Engagement Scale. Figure 4.6 presents the SES CCC. The figure shows that all five response categories have the highest probability of being selected at different points along the construct continuum. This is encouraging; however, the area of difference between person and item estimate where category 3 (*about the same as X*) is most likely to be chosen is quite small. This provides further evidence that the items should be revised.

The intersections of each response category curve represent the Andrich thresholds, which were also illustrated in the variable map presented in Figure 4.4. These thresholds, their fit statistics, and the average logit estimates of individuals making using of each category, are presented in Table 4.6. Unsurprisingly, the threshold is lowest for the transition between categories 1 and 2, and is highest for the transition between categories 4 and 5. The average estimates are also in intuitive order; response category 1 has the lowest average estimate and response category 5 has the highest. These numerical indicators corroborate what is shown visually in the category characteristic curve. Response category 1 has both INFIT and OUTFIT statistics that are slightly above the 1.3 criterion, indicating the possibility of both inconsistent and greatly unexpected responses. The low frequency of use for category 1 may contribute to these fit statistics.

Figure 4.6

Social Engagement Scale Category Characteristic Curve (Pre-pilot)

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+------+------+------+------+------+------+------+------+-
R  1.0 +                                                         +
O      |                                                         |
B      |                                                         |
A      |11                                                     55|
B   .8 +  11                                                 55  +
I      |    11                                             55    |
L      |      1                                          55      |
I      |       11                                       5        |
T   .6 +         1                                     5         +
Y      |          1    2222222              44444    55          |
    .5 +           1122       22          44     44 5            +
O      |           221          2       44         *4            |
F   .4 +          2   1          22    4          5  44          +
       |        22     1         33*33*3        55     44        |
R      |      22        11     33   *4  33     5         4       |
E      |    22            1  33    4 2    33  5           44     |
S   .2 +  22               **     4   22    **              444  +
P      |22               33  11 44      2 55  33               44|
O      |              333     4*1       5*2     333              |
N      |         33333    4444   111*555   2222    33333         |
S   .0 +******************5555555555 1111111111******************+
E      -+------+------+------+------+------+------+------+------+-
       -4     -3     -2     -1      0      1      2      3      4
        PERSON [MINUS] ITEM MEASURE
```

Table 4.6

Social Engagement Scale Andrich Thresholds and Average Estimates (Pre-pilot)

Response Category	Andrich Threshold	Response Frequency	INFIT	OUTFIT	Average Estimates
1 (X is much more engaged)	N/A	6	1.31	1.36	-.50
2 (X is a little more engaged)	-2.24	36	1.10	1.08	-.06
3 (about the same as X)	-.19	57	.94	.83	.68
4 (a little more engaged than X)	.25	129	1.21	.94	1.37
5 (much more engaged than X)	2.18	140	.93	.93	2.58

Residual Analysis

The final piece of analysis for each of the OCAS was a principal components analysis (PCA) on the Rasch residuals. This process checks for evidence of an uncaptured construct dimension within the residuals. Unlike a traditional PCA, where one typically seeks evidence of a strong underlying component, the ideal outcome of this analysis is to find no relationship among the residuals. The results of the residual PCA can be compared to randomly generated data; ideally, these results would be similar.

Physical Accessibility Scale. Table 4.7 presents the eigenvalues and percentages of variance explained obtained from the PCA of the PAS residuals alongside eigenvalues obtained from a PCA of randomly generated data. While not perfect, the results are encouraging. Although a component with an eigenvalue greater than 2 was extracted from the PAS residuals, the difference in variance explained by this component and the first component extracted from the random data is less than 10%. This finding is complemented by the scree plots for the PAS residuals and random data, which are presented in Figures 4.7 and 4.8, respectively. Although the first extracted component again stands out among the PAS residuals, there is no visual evidence of a sharp break that would indicate a clear unidimensional PCA solution. The component loading plot provides a final piece of evidence concerning the randomness of the PAS residuals (Figure 4.9). While the residuals are not exactly a perfectly circular pattern (which would indicate that they are truly random), they do approximate a circular pattern and are not clustered around the first two extracted components.

Table 4.7

Principal Components Analysis Results for PAS Residuals and Random Data (Pre-pilot)

PAS Residuals			Randomly Generated Data		
Component Number	Eigenvalue	% Variance Explained	Component Number	Eigenvalue	% Variance Explained
1	2.701	30.008	1	1.898	21.088
2	1.585	17.610	2	1.493	16.586
3	1.232	13.693	3	1.236	13.739
4	1.024	11.381	4	1.082	12.019
5	.860	9.558	5	1.024	11.375
6	.730	8.116	6	.809	8.993
7	.511	5.675	7	.688	7.648
8	.339	3.765	8	.430	4.776
9	.017	.193	9	.340	3.777

Figure 4.7

Scree Plot for PAS Residuals (Pre-pilot)

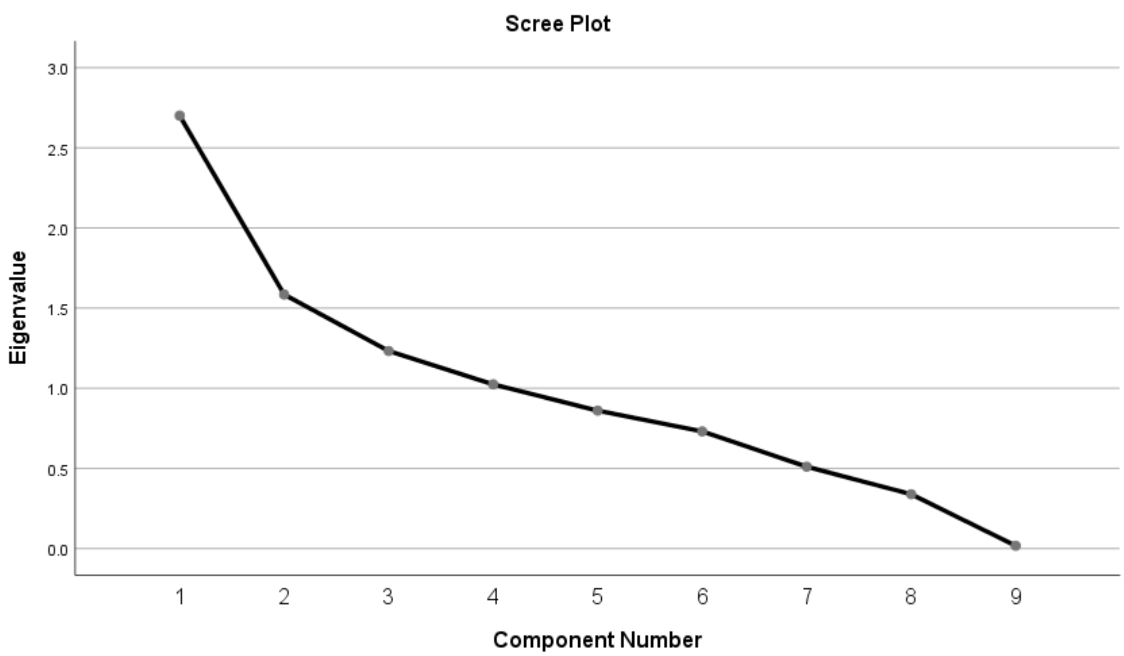


Figure 4.8

Scree Plot for Randomly Generated Data (PAS Comparison)

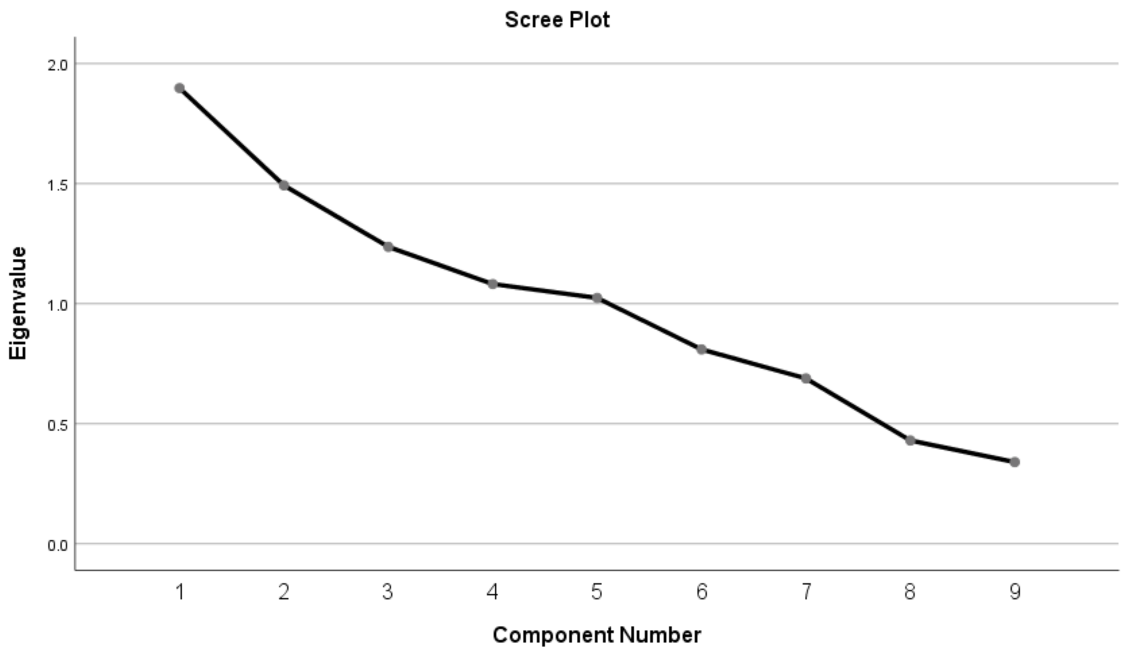
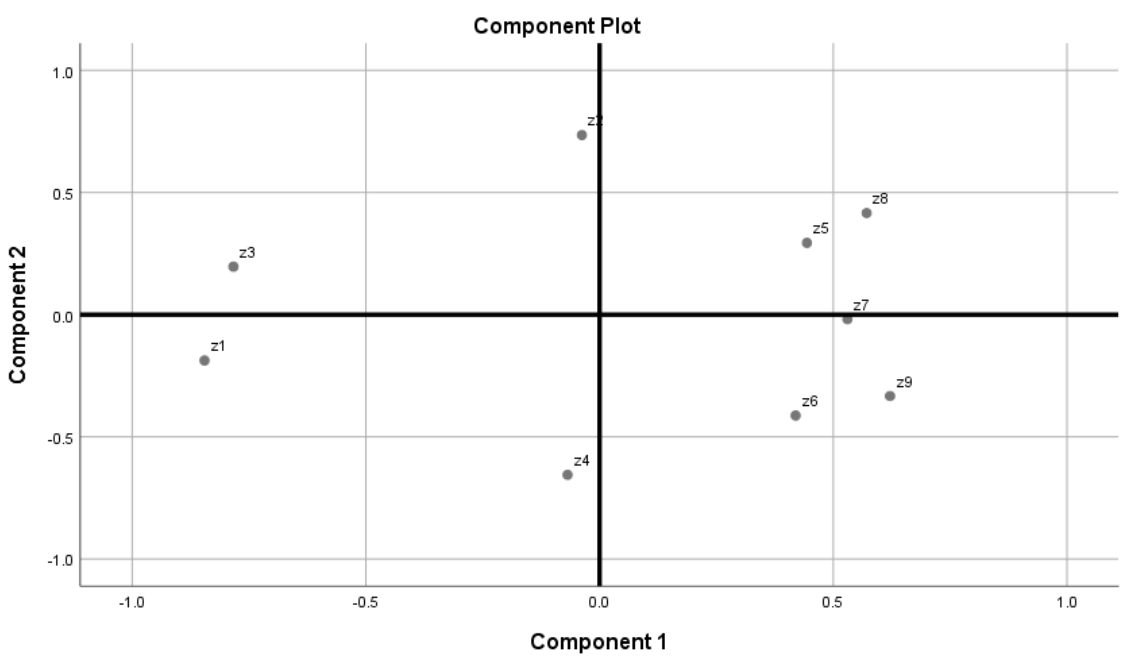


Figure 4.9

Physical Accessibility Scale Residual Component Loading Plot (Pre-pilot)



Social Engagement Scale. Table 4.8 presents the eigenvalues and percentages of variance explained obtained from the PCA of the SES residuals with eigenvalues obtained from a PCA of randomly generated data. These results are also encouraging. Although a component with an eigenvalue greater than 2 was extracted from the SES residuals, the difference in variance explained by this component and the first component extracted from the random data is less than 10%. The scree plots presented in Figures 4.10 and 4.11 show that the first extracted component among the residuals does stand out visually compared to that of the random data, however, this is not a clear unidimensional solution. The component loading plot for the SES appears less random than the PAS; instead of approximating a circular pattern, the residuals are clustered in two groups (see Figure 4.12). The three items reflecting the highest levels of social engagement are in one group, while all other items are in a second group. This provides further evidence that the SES items could benefit from revision.

Table 4.8

Principal Components Analysis Results for SES Residuals and Random Data (Pre-pilot)

SES Residuals			Randomly Generated Data		
Component Number	Eigenvalue	% Variance Explained	Component Number	Eigenvalue	% Variance Explained
1	2.432	27.020	1	1.934	21.484
2	1.371	15.234	2	1.571	17.459
3	1.261	14.015	3	1.262	14.021
4	1.050	11.671	4	1.109	12.318
5	.888	9.864	5	.969	10.768
6	.847	9.417	6	.777	8.634
7	.658	7.310	7	.638	7.085
8	.474	5.264	8	.417	4.638
9	.019	.207	9	.323	3.592

Figure 4.10

Scree Plot for SES Residuals (Pre-pilot)

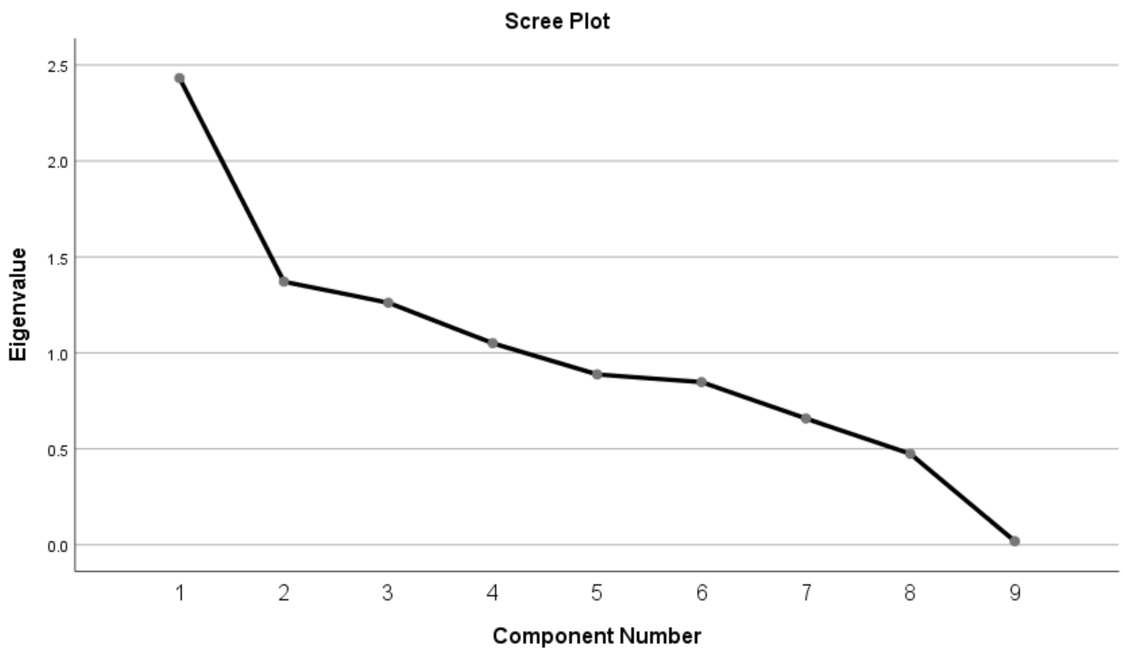


Figure 4.11

Scree Plot for Randomly Generated Data (SES Comparison)

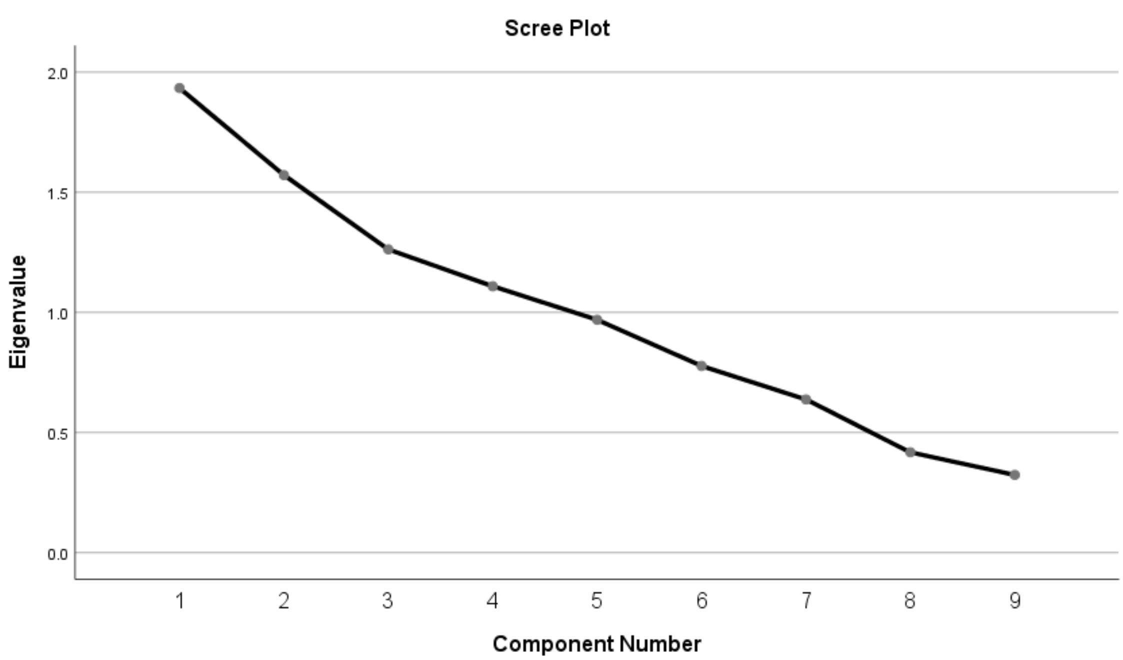
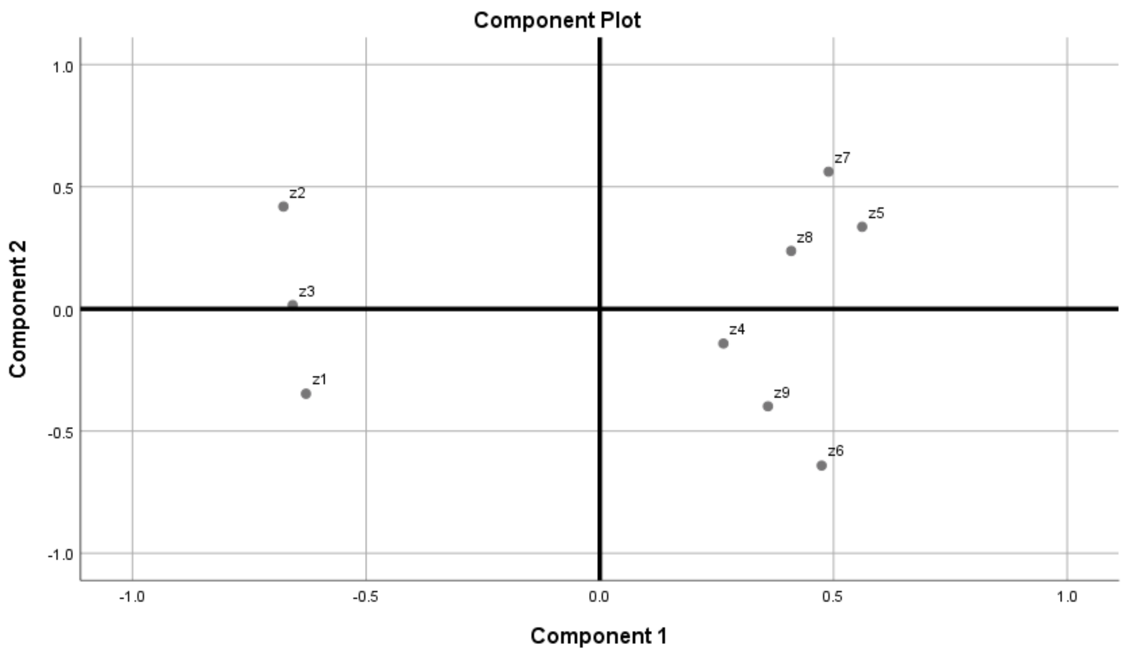


Figure 4.12

Social Engagement Scale Residual Component Loading Plot (Pre-pilot)



Item Revisions

Based on the results of the analyses presented above, revisions were made to both the PAS and SES items. For revised items, Tables 4.9 and 4.10 present the pre-pilot version of the item, the revised item for the pilot study, and the rationale for the changes. Changes to the items are bolded in the pilot versions. More changes were made to the SES than the PAS. The PAS did not display any major issues with disordered items and it was hypothesized with minor adjustments and a larger sample size that better spread might be obtained for the items. However, the SES had one item (MMM) that was severely disordered and required substantial revision.

Table 4.9

Physical Accessibility Scale Item Revisions Based on Pre-pilot

Pre-pilot Item	Pilot Item	Rationale for Changes
Professor Mason's regular office hours are not sufficient for the number of meeting requests she receives; students sometimes must wait over a week to meet with her. She sometimes crosses paths with students around campus and usually responds to emails in a timely manner, but students sometimes do not receive a reply for several days.	Professor Mason **cannot always quickly accommodate requests she receives for meetings**; students sometimes must wait over a week to meet with her. She sometimes crosses paths with students around campus and usually responds to emails in a timely manner, but students sometimes do not receive a reply for several days.	The Arranged Meetings facet for this item was adjusted to make the statement slightly more positive; it is intended to convey that Professor Mason does have a decent number of office hours; however, they are not always sufficient to accommodate all student requests. I was concerned that the pre-pilot version conveyed that Professor Mason simply did not have enough office hours.
Professor Johnson does not have scheduled office hours and typically does not accommodate student requests for meetings; students rarely see her outside of class at all. However, she does usually respond to student emails, albeit not always in a timely manner.	Professor Johnson does not have scheduled office hours and typically does not accommodate student requests for meetings; students rarely see her outside of class at all. **She does respond to student emails, but not always in a timely manner.**	The Email facet for this item was revised to make the wording clearer.

Table 4.10

Social Engagement Scale Item Revisions Based on Pre-pilot

Pre-pilot Item	Pilot Item	Rationale for Changes
Professor Bradley is clearly engaged during meetings with students; he listens carefully to student concerns or questions and engages in thoughtful discussion. When he encounters students around campus, Professor Bradley gives enthusiastic greetings and seems happy to see them. His responses to student emails are always courteous and respectful.	Professor Bradley is clearly engaged during meetings with students; **he listens to students carefully, asks questions, and engages in thoughtful discussion. He seems happy to see students around campus and often stops to chat with them.** Professor Bradley's responses to emails are always courteous and **clearly address students' questions or concerns.**	Given that 4 respondents selected the top response category for all SES items, this item was revised to be more difficult overall. The additional piece of "asking questions" was added to the Arranged Meetings facet. The Chance Encounters facet was revised for language clarity. The Email facet was revised to include a description of the clarity of Professor Bradley's emails.
Professor Jenkins clearly pays attention to students during meetings; she does not seem at all distracted by other matters. When students see Professor Jenkins in the halls, she typically offers a friendly greeting. Her responses to emails are sometimes abrupt, but never rude or disrespectful.	Professor Jenkins clearly **prioritizes** students during meetings and does not seem at all distracted by other matters. **She often greets students when she encounters them around campus.** Her responses to emails are sometimes abrupt, but never rude or disrespectful **and usually adequately address students' concerns or questions.**	The Arranged Meetings facet was adjusted to be more difficult—"prioritizes" is a stronger statement than "pays attention to". The Chance Encounters facet was revised for language clarity. The Email facet was revised to include a description of the clarity of Professor Jenkins' emails.
Professor Howard prioritizes and cares about meetings with students; he does not make students feel as though they are intruding. His responses to emails also address student concerns in a profession tone. However, he does not consistently acknowledge students when he sees them around campus.	Professor Howard **cares about meetings with students; he is attentive** and does not make students feel as though they are intruding. His **responses to emails address** student concerns in a professional tone. However, he does not consistently acknowledge students when he sees them around campus.	The Arranged Meetings facet was revised to vary the word in comparison with the scenario above. "Also" was removed from the Chance Encounters facet because it was deemed unnecessary.

Professor Thomas listens to students' concerns and questions during meetings, but does not always engage in thoughtful conversation. She sometimes greets students when they cross paths around campus, but her tone seems insincere. Professor Thomas's responses to students' emails, however, are thorough and respectful.	Professor Thomas listens to students' concerns and questions during meetings, but does not always engage in thoughtful conversation. **She acknowledges students inconsistently when they cross paths around campus.** Professor Thomas's responses to students' emails, however, are thorough and respectful.	The Chance Encounters facet was revised for clarity and to lessen its harshness; the intent of the facet is to capture whether or not Professor Thomas socially engages with students, not for students to speculate about her sincerity.
Professor Medina occasionally seems distracted during meetings, making students feel as though they are bothering him. He may acknowledge students when he sees them around campus, but rarely starts up a conversation. Professor Medina's responses to emails are courteous, although he does not always clearly address students' questions and concerns.	**Professor Medina occasionally seems distracted during meetings with students.** He may acknowledge students when he sees them around campus, but rarely starts up a conversation. Professor Medina's responses to emails are courteous, although he does not always clearly address students' questions and concerns.	This item had a severe disordering in the pre-pilot. It was hypothesized that the language of "making students feel as though they are bothering him" in the Arranged Meetings facet was interpreted much more harshly by respondents than intended; this language was removed for the pilot.
Professor Larson is generally focused during meetings with students, although she may not always engage in thoughtful conversations. Her responses to emails are not always helpful, but are generally courteous. Professor Larson rarely acknowledges students when she encounters them outside of class or office hours.	Professor Larson is generally focused during meetings with students, although she may not always engage in thoughtful conversations. **Her responses to student emails are not always clear, but are always respectful.** Professor Larson rarely acknowledges students when she encounters them outside of class.	The language in the Email facet was adjusted to reduce the influence of respondents' perceptions of "helpfulness"; the sentence is intended to capture Professor Larson's engagement with students via email, not necessarily how helpful they perceived her to be.
Professor Taylor is often preoccupied with other tasks during meetings with students. She also does not go out of her way to greet students around	Professor Taylor is often preoccupied with other tasks during meetings with students. **She does not go out of her way** to greet students around campus	"Also" was removed from the Chance Encounters facet because it was deemed unnecessary.

campus and seems annoyed if students try to initiate conversation. However, Professor Taylor’s responses to emails are generally clear and professional.	and seems annoyed if students try to initiate conversation. However, Professor Taylor’s responses to emails are generally clear and professional.	

Pilot Study

Overview of Responses

1,500 Boston College undergraduate students were invited to complete the pilot survey. These students were sampled randomly using a list obtained from Boston College's Office of Institutional Research and Planning. The pilot instrument was administered using Qualtrics and data were collected over a 6-week period during March and April 2019. The purpose of the pilot study was to serve as another empirical check on the OCAS' conception of the physical accessibility and social engagement constructs, as well as to highlight any items requiring additional revisions prior to final administration. Extensive detail regarding Rasch output is not presented for the pilot study; this information can be reviewed in the pre-pilot results.

Missing Data

Qualtrics captured 130 submitted surveys for the pilot study. Of these 130 records, 64 (49.2%) were completely blank (i.e., no item responses were recorded). These 64 records were excluded at the outset of pilot analyses, yielding a starting sample size of 66. Of the 66 records containing data, 52 (80.3%) were complete (i.e., all nine items for both scales had recorded responses). For the fourteen cases that were not complete, ten (15.1% of total) records contained complete data for the Physical Accessibility Scale (PAS), but no data for the Social Engagement Scale (SES). For the remaining four cases, one individual skipped a single SES item, one individual skipped all but one of the SES items, one individual responded to only 3 PAS items, and one individual responded to only a single PAS item. A summary of missing data from the pilot administration is provided in Table 4.11.

Table 4.11

Missing Data Summary (Pilot)

Physical Accessibility Scale			
Number of Missing Variables	Frequency (Number of Cases)	Percent	Cumulative Percent
8	1	1.5	1.5
5	1	1.5	3.0
0	64	97.0	100.0
Social Engagement Scale			
Number of Missing Variables	Frequency (Number of Cases)	Percent	Cumulative Percent
9	12	18.1	18.1
8	1	1.5	19.6
1	1	1.5	21.1
0	52	78.9	100.0

This is a manageable amount of missing data. The twelve respondents who completed only the PAS items most likely quit the Qualtrics survey after completing the first screen, as the two scales were presented on two different pages. These individuals were retained for the Rasch analysis of the PAS, but excluded from the SES. Individuals who only completed a single item for each scale were excluded from all analyses, as they did not provide a sufficient amount of information for the scales to adequately capture their perceptions. Finally, the individual who completed only three PAS items was also excluded as they completed fewer than half of the scale items. This leaves the two scales with different sample sizes for the pilot analyses: 64 individuals for the PAS (4.2% response rate) and 53 individuals for the SES (3.5% response rate). One respondent selected 5 (the highest category) for all 9 items on the PAS and SES, so they are automatically excluded from the Rasch analyses but are included in the descriptive statistics.

One additional exclusion was made from the PAS pilot sample upon examination of individual response patterns in conjunction with the Rasch analysis results. This respondent exhibited poor fit statistics (INFIT of 7.83; OUTFIT of 9.90) and their response pattern indicated a possibility that they interpreted the response options in reverse order. Examination of the PAS Rasch results including and excluding this individual revealed that they were exerting some influence on the fit statistics for the scale items. Because the goal of this study is to provide support for the instrument calibration rather than to obtain precise estimates for individuals, this person is excluded from the following descriptive and Rasch results. PAS Rasch analysis results that do include this individual are presented in Appendix I for comparison.

A single respondent skipped one item within SES. This is not suggestive of a pattern. This individual was retained for the pilot Rasch analyses because the estimation procedure is capable of addressing missing data.

Descriptive Statistics

Respondents

Although pilot study respondents were not asked to provide demographics, this information was obtained through the records of Boston College's institutional research office. This demographic information is not associated with survey responses in order to protect student privacy. This does limit my ability to diagnose differential responding among student groups; however separation of demographic information from item responses is a condition of my IRB approval. Table 4.12 provides information regarding respondents' gender, race/ethnicity, and the college in which they are enrolled. Race/ethnicity categories are those used in the Integrated Postsecondary Education Data

System (IPEDS). Demographics are presented only for the 63 respondents who are included in the Rasch analyses. Demographic data are presented for all 1,500 invitees in Table 4.13.

Table 4.12

Pilot Study Respondents' Demographic Information

	Number of Respondents	Percentage of Respondents
Gender		
Male	17	27.0
Female	46	73.0
Race/Ethnicity		
Asian	11	17.5
Black or African American	1	1.6
Hispanic of Any Race	5	7.9
International	7	11.1
Race/Ethnicity Unknown	2	3.2
Two or More Races	2	3.2
White	35	55.6
College		
Arts and Sciences	36	57.1
Education	8	12.7
Management	15	23.8
Nursing	4	6.3

Table 4.13

Pilot Study Invitees' Demographic Information

	% of Invitees
Gender	
Male	46.3
Female	53.7
Race/Ethnicity	
Asian	9.3
Black or African American	4.8
Hispanic of Any Race	11.5
International	7.9
Race/Ethnicity Unknown	4.0
Two or More Races	3.7
White	58.8
College	
Arts and Sciences	65.6
Education	6.5
Management	23.3
Nursing	4.6

Scenario Items

Table 4.14 presents the means and standard deviations for each of the PAS and SES scenario items. Items are presented within the table in their hypothesized order from most difficult (high level across all three facets) to least difficult (low level across all three facets). Item naming conventions are identical to the pre-pilot: "H" indicates a high facet level, "M" indicates a medium facet level, and "L" indicates a low facet level. Facets are always listed in the following order: arranged meetings, chance encounters, email. The number in parentheses indicates the overall "level" of the item as calculated by summing the individual facet levels. Although the items are numbered within the table, this is for reference purposes only; items were randomized when presented to respondents.

For the PAS, the item with the highest mean is Item 9 (LLL). Item 9 (LLL) also has the highest mean for the SES. The coding scheme for response options is the same for the pilot study as it was for the pre-pilot, which means that the high means for these particular items suggest that they were the easiest items for respondents to select the highest level option of *much more accessible/engaged than X*. In contrast, the items with the lowest means are those for which it was most difficult for respondents to select the highest level option of *much more accessible/engaged than X*.

Examination of the item means in conjunction with their designated facet levels provides a preview of what will be confirmed by the Rasch analysis. As in the pre-pilot, items with higher facet levels tend to have lower means, meaning that it is more difficult for respondents to select the higher response categories that would indicate the faculty member they had in mind while completing the items is more accessible/engaged than the fictitious faculty member in the item scenario. The items constructed to reflect low facet levels tend to have higher means, meaning that it is easier for respondents to select the higher response categories. This provides some preliminary evidence that the hypothesized construct continua for the PAS and SES function as intended. Revisions made to SES items between the pre-pilot and pilot administrations appear to have been successful; while Items 5 (MMM) and 6 (MLM) are reversed, the other scale item means all fall in their intended order.

Table 4.14

Descriptive Statistics for Scenario Items (Pilot)

Physical Accessibility Scale		
	Mean	Standard Deviation
Item 1: HHH(9)	2.49	0.84
Item 2: HHM (8)	3.73	1.05
Item 3: HMH (8)	2.78	1.04
Item 4: MMH (7)	3.81	0.82
Item 5: MMM (6)	4.44	0.67
Item 6: MLM (5)	4.33	0.93
Item 7: MML (5)	4.49	0.82
Item 8: LLM (4)	4.79	0.48
Item 9: LLL (3)	4.76	0.56
Social Engagement Scale		
	Mean	Standard Deviation
Item 1: HHH(9)	2.66	0.90
Item 2: HHM (8)	3.42	0.82
Item 3: HMH (8)	3.66	0.81
Item 4: MMH (7)	3.89	0.93
Item 5: MMM (6)	4.26	0.74
Item 6: MLM (5)	4.15	0.83
Item 7: MML (5)	4.43	0.92
Item 8: LLM (4)	4.45	0.85
Item 9: LLL (3)	4.83	0.43

Table 4.14 provides some preliminary evidence of effectiveness for revisions made to the pre-pilot items. One aim for both scales was to increase the overall difficulty of items. Item means from the pilot administration are generally lower than their pre-pilot counterparts. The item means are also more spread out in the pilot results compared to the pre-pilot. For the PAS, the pre-pilot item means ranged from 3.28 to 4.53 (range of 1.25); for the pilot these item means ranged from 2.49 to 4.79 (range of 2.30). Similarly, for the SES, the pre-pilot item means ranged from 3.22 to 4.46 (range of 1.24); for the pilot, these item means ranged from 2.66 to 4.83 (range of 2.17). This provides additional evidence that the functioning of the scale has improved from the pre-pilot study.

Rasch Analyses

As with the pre-pilot, I performed a Rasch rating scale analysis (Andrich, 1978) on the pilot data. This round of analysis had three purposes: 1) to provide further confirmation of the hypothesized construct continua that were initially supported by the pre-pilot; 2) to evaluate the item revisions made following the pre-pilot; and 3) to determine if further revisions were needed prior to the instrument's final administration. Pilot study Rasch analysis results are presented in the same order and fashion as for the pre-pilot.

Variable Maps

Physical Accessibility Scale. Figure 4.13 presents the PAS variable map. The line running up the center of the map represents the hierarchical continuum physical accessibility outside of class. Higher representations and perceptions of physical accessibility at the top of the map and lower representations and perceptions of physical accessibility are at the bottom. Markers for persons and items along the variable are the same as what was presented in the pre-pilot results.

Figure 4.13

Physical Accessibility Scale Variable Map (Pilot)

```
MEASURE                        PERSON - MAP - ITEM
                                   <more>|<rare>
    7                               164  +
                                         |
                                         |
                                         |
    6                                    +
                                         |
                                         |
                                        T|
    5                          154  158  +
                                         |
                                         |
           132  134  149  152  159  160  |T
    4                                    +
      127  133  136  137  147  150  156 S|
           109  110  135  138  151  155  |  HHH (9)
                                         |
    3           115  142  143  161  162  +  HMH (8)
                     111  123  144  157  |
                                        M|
                130  131  146  148  153  |
    2      103  114  120  124  125  128  +S
                     104  105  119  163  |
                               121  140  |
                               108  129  |
    1                          117  126 S+  HHM (8)
                               141  145  |  MMH (7)
                               116  139  |
                                    101  |
    0                          112  118  +M
                                    107  |
                               102  113 T|
                                    106  |  MLM (5)
   -1                                    +  MMM (6)
                                         |  MML (5)
                                         |
                                         |
   -2                                    +S
                                         |
                                         |  LLL (3)
                                         |  LLM (4)
   -3                                    +
                                   <less>|<freq>
```

The items generally follow the hypothesized progression from low levels of perceived physical accessibility to high levels of perceived physical accessibility. This variable map also highlights improvements from the pre-pilot study. For example, Item HHH is now clearly at the top of the map on the item side, whereas it was side by side with Item HMH in the pre-pilot. This indicates that Item HHH was most difficult for respondents to select the highest level option of *much more accessible/engaged than X*.

Despite this improvement from the pre-pilot at the high end of the scale, the PAS pilot results deviate somewhat from what was expected at the low end of the scale. Item LLL is near the bottom of the variable map, but Item LLM is the easiest item overall. Items in the middle of the scale are also clustered close together. These issues suggest that the scale might benefit from additional revisions to better differentiate the scenarios.

Although the descriptive statistics and variable map reveal that the PAS items were more difficult in the pilot study, there is still some degree of mismatch between persons and items. This is illustrated by the difference between the average person estimate (2.47) and the average item estimate (set to 0.0). However, there is still clear improvement from the pre-pilot results. In the pre-pilot, a substantial portion of respondents fell above the most difficult item on the PAS variable map, meaning that it was not possible to describe what a very high perception of physical accessibility “looks like” because there are no items in that area of the scale. This number of individuals has decreased in the pilot study.

The person separation and reliability for the PAS pilot are 2.10 and .82, respectively. This is another improvement from the pre-pilot, which had a separation of 1.81 and .77. The increase in person separation and reliability indicate that the pilot PAS items are better able to differentiate among individuals with high and low perceptions of physical accessibility. However, there is still room for additional improvement. The separation and reliability also improved for items between the pilot and pre-pilot studies

(7.87 versus 3.37 and .98 versus .92). However, these statistics are not comparable because of the differences in sample size between the two administrations.

Figure 4.14

Physical Accessibility Scale Variable Map with Andrich Thresholds (Pilot)

```
MEASURE  PERSON - MAP - ITEM - Andrich thresholds (modal categories if ordered)
         <more>|<rare> 2       3       4       5
    7       X  +
               |
               |                                            HHH     .5
               |
    6          +
               |                                            HMH     .5
               |
              T|
    5      XX  +
               |
               |                              HHH     .4
   XXXXXX  |T
    4          +
  XXXXXXX S|                              HMH     .4    HHM     .5
                                                            MMH     .5
   XXXXXX  |
               |
    3   XXXXX  +
         XXXX  |                MHH     .3
              M|
        XXXXX  |                HMH     .3                  MLM     .5
    2  XXXXXX  +S                                           MMM     .5
         XXXX  |                              HHM     .4    MML     .5
           XX  |                              MMH     .4
           XX  |
    1      XX S+
           XX  |
           XX  |    HHH     .2                              LLL     .5
            X  |                HHM     .3    MLM     .4    LLM     .5
    0      XX  +M               MMH     .3
            X  |    HMH     .2                MMM     .4
                                              MML     .4
           XX T|
            X  |
   -1          +
               |
               |                MLM     .3    LLL     .4
               |                MMM     .3    LLM     .4
   -2          +S               MML     .3
               |    HHM     .2
                    MMH     .2
               |
               |
   -3          +
               |                LLL     .3
               |                LLM     .3
               |    MLM     .2
   -4          +    MMM     .2
               |T   MML     .2
               |
               |
   -5          +
               |
               |    LLL     .2
               |    LLM     .2
   -6          +
         <less>|<freq> 1       2       3       4
```

Figure 4.14 shows a variable map that includes the Andrich thresholds, which represent the point at which an individual has a 50% probability of selecting the next higher response category for the item. In this map, each item appears four times: once for each response category transition. The item labels followed by ".2" are the Andrich thresholds for moving between response categories one and two (*X is much more accessible* and *X is a little more accessible*), labels followed by ".3" are the Andrich thresholds for moving between response categories two and three (*X is a little more accessible* and *about the same as X*), and so on. Respondents are represented by "X" and the placement of individuals relative to items indicates which response categories they were likely to select.

Figure 4.14 also illustrates how many individuals in the sample were relatively high-scoring on the PAS. Most respondents were expected to select at least *about the same as X* (score of 3) for all items. This is also an improvement from the pre-pilot, where most respondents were expected to select at least *a little more accessible than X* for most items. This finding is complementary to the item means and the first variable map; while individuals were relatively high scoring overall, the item spread has improved between the pre-pilot and pilot study.

Social Engagement Scale. Figure 4.15 presents the SES variable map. The line in the center of the map represents the hierarchical continuum of students' perceptions of a faculty member's social engagement outside of class. The layout of this variable is comparable to the PAS: higher perceptions of social engagement are represented at the top of the map and lower perceptions of social engagement are represented at the bottom.

Figure 4.15

Social Engagement Scale Variable Map (Pilot)

```
MEASURE                        PERSON - MAP - ITEM
                                   <more>|<rare>
    7                               163  +
                                         |
                                         |
                                         |
    6                                    +
                                         |
                                         |
                                    109  |
    5                                   T+
                                         |
                          126  133  159  |
                                         |
    4                151  155  158  161  +
      110  113  146  148  153  157  160 S|
                                         |T
                               108  154  |  HHH (9)
    3                               141  +
                                         |
           103  105  106  115  132  139  |
                     118  127  142  147 M|
    2                107  135  145  150  +
                          114  123  130  |S
                     104  120  125  137  |  HHM (8)
                               119  129  |
    1                     101  122  156 S+  HMH (8)
                                    112  |
                               102  138  |  MMH (7)
                               117  140  |
    0                                    +M
                          111  128  162  |  MLM (5)
                                        T|  MMM (6)
                                         |
   -1                                    +  MML (5)
                                         |  LLM (4)
                                         |
                                         |S
   -2                                    +
                                         |
                                         |
                                         |
   -3                                    +  LLL (3)
                                   <less>|<freq>
```

The SES variable map has improved dramatically from the pre-pilot. Item MMM, which had been at the very bottom of the map, is now much closer to its intended position. The spread of items across the scale could still be increased; however, the items are much more spread out in the pilot results than they were in the pre-pilot.

As with the PAS pilot results, on average, the item estimates are below the person estimates. The average person estimate is 2.42 and the average item estimate is set to 0.0.

This is also apparent in the distribution of individuals and items along the variable map. However, the pilot study had only a single individual with a “perfect” SES score, compared to five individuals in the pre-pilot.

Person and item separation also improved for the SES between the pre-pilot and pilot studies. The person separation and reliability increased modestly from 1.93 and .79 to 2.00 and .80. This suggests that high- and low-scoring individuals are better differentiated in the pilot study than they were in the pre-pilot; however, there is still room for improvement. Increases were more pronounced for items, where the separation increased from 2.88 to 6.18 and the reliability increased from .89 to .97; however, this is solely due to the increase in sample size and does not necessarily imply improvement in construct definition.

Figure 4.16

Social Engagement Scale Variable Map with Andrich Thresholds (Pilot)

```
MEASURE  PERSON - MAP - ITEM - Andrich thresholds (modal categories if ordered)
         <more>|<rare> 2        3        4        5
    7       X  +
               |
               |
               |                                              HHH     .5
    6          +
               |
               |
            X  |
    5         T+
               |
          XXX  |                                HHH     .4    HHM     .5
               |
    4    XXXX  +                                              HMH     .5
      XXXXXXX S|
               |T                                             MMH     .5
           XX  |
    3       X  +
               |                                HHM     .4    MLM     .5
       XXXXXX  |                                              MMM     .5
         XXXX M|                                HMH     .4
    2    XXXX  +                                              MML     .5
          XXX  |S                 HHH     .3    MMH     .4    LLM     .5
         XXXX  |
           XX  |
    1     XXX S+                                MLM     .4
            X  |                                MMM     .4
           XX  |    HHH     .2
           XX  |                                MML     .4
    0          +M                 HHM     .3    LLM     .4    LLL     .5
          XXX  |
              T|                  HMH     .3
               |
   -1          +                  MMH     .3
               |    HHM     .2
               |
               |S   HMH     .2    MLM     .3    LLL     .4
   -2          +                  MMM     .3
               |    MMH     .2
               |                  MML     .3
               |                  LLM     .3
   -3          +    MLM     .2
               |    MMM     .2
               |T
               |    MML     .2
   -4          +    LLM     .2
               |
               |                  LLL     .3
               |
   -5          +
               |
               |
               |    LLL     .2
   -6          +
         <less>|<freq> 1        2        3        4
```

Figure 4.16 shows a variable map that includes the Andrich thresholds, which represent the point at which an individual has a 50% probability of selecting the next higher response category for the item. In this map, each item appears four times: once for each response category transition. The numbers following each item label represent the transition point to that particular response category, as in Figure 4.14. Also comparable to Figure 4.14, individuals are represented by "X".

Most individuals were expected to select *a little more engaged than X* for most items. This complements the finding that respondents generally had high scores on the SES. This version of the variable map also highlights the improved spread of the items compared to the pre-pilot.

Fit Statistics

As with the pre-pilot, Rasch model fit statistics are used in the pilot to flag potentially problematic individuals or items. Again, item fit statistics for items are of primary concern as the goal at this stage is not to provide precise estimates for individuals.

Physical Accessibility Scale. Table 4.15 presents the fit statistics for the PAS pilot data. The items are listed in difficulty order, as indicated in the logit estimate column. A higher logit estimate indicates an item for which it is more difficult for respondents to select the upper-level categories (i.e., to indicate that the faculty member they have in mind is at least as accessible as the fictitious faculty member in the item).

Table 4.15

Physical Accessibility Scale Fit Statistics (Pilot)

Item	Logit Estimate (S.E.)	Information-Weighted Fit Statistic		Unweighted Fit Statistic	
		MNSQ	ZSTD	MNSQ	ZSTD
Item 1: HHH (9)	3.51 (.19)	1.01	.14	1.09	.54
Item 3: HHM (8)	2.90 (.18)	.88	-.65	.86	-.83
Item 2: HMH (8)	.89 (.19)	1.14	.84	1.15	.87
Item 4: MMH (7)	.71 (.19)	.74	-1.61	.72	-1.72
Item 6: MLM (5)	-.66 (.22)	1.04	.26	.87	-.57
Item 5: MMM (6)	-1.02 (.24)	1.07	.42	.95	-.12
Item 7: MML (5)	-1.19 (.24)	1.02	.16	.73	-1.04
Item 9: LLL (3)	-2.47 (.32)	1.58	1.95	1.85	1.57
Item 8: LLM (4)	-2.68 (.34)	1.17	.67	1.07	.31

Eight of the nine scale items have acceptable fit statistics, which is an improvement from the pre-pilot. These statistics indicate that the items are generally a good fit to the Rasch model. Item 9 (LLL), which had INFIT and OUTFIT statistics greater than 1.3, is highlighted in Table 4.14. It is also worth noting that this particular item did not fall in its hypothesized place along the variable map continuum. Given that it was constructed to reflect low levels of accessibility across all three facets, it was hypothesized to be the easiest item; however, this did not turn out to be the case.

I examined the person-response table to further evaluate Item 9. The person-response table for the pilot can be found in Appendix J. Examination of this table revealed one individual who scored unexpectedly low on Item 9; this particular response was so unexpected that it caused the individual to have a printed OUTFIT statistic of 9.9. This was by far the most unexpected response for Item 9, so I removed this individual from the sample and re-ran the Rasch analysis. When this person was removed, not only were the problematic fit statistics for Item 9 resolved, but Item 9 also fell in its intended position at the bottom of the variable map as the easiest item. These findings lend support to the hypothesized construct continuum and item fit to the Rasch model, as the disordering and problematic fit statistic for Item 9 were attributable to a specific individual and not necessarily a problem with the hypothesized construct or items. Despite this individual's influence on the variable map and item fit statistics, they were retained as a part of the pilot sample because there were no clear issues with the rest of their responses (unlike the exclusion discussed in the Overview of Responses section above).

Social Engagement Scale. Table 4.16 presents the fit statistics for the SES pilot data. Items again appear in Rasch difficulty order.

Table 4.16

Social Engagement Scale Fit Statistics (Pilot)

Item	Logit Estimate (S.E.)	Information-Weighted Fit Statistic		Unweighted Fit Statistic	
		MNSQ	ZSTD	MNSQ	ZSTD
Item 1: HHH (9)	3.20 (.20)	1.76	3.34	1.79	3.37
Item 2: HHM (8)	1.53 (.21)	.77	-1.23	.77	-1.22
Item 3: HMH (8)	.96 (.21)	.70	-1.70	.70	-1.68
Item 4: MMH (7)	.41 (.22)	.84	-.80	.81	-.97
Item 6: MLM (5)	-.24 (.23)	.59	-2.42	.53	-2.55
Item 5: MMM (6)	-.60 (.24)	.65	-1.93	.63	-1.75
Item 7: MML (5)	-1.07 (.25)	1.72	2.96	1.28	1.03
Item 8: LLM (4)	-1.20 (.25)	1.34	1.54	1.05	.26
Item 9: LLL (3)	-2.99 (.38)	.94	-.11	2.24	1.60

Six of the nine scale items have acceptable fit statistics. Items 1 (HHH), 7 (MML), and 8 (LLM) all have INFIT statistics greater than 1.3; Item 1 also has an

OUTFIT statistic greater than 1.3. Generally, these statistics suggest that the source of misfit is a general inconsistency of responses rather than individual surprising responses.

I examined the person-response table (see Appendix J) to further evaluate the three items exhibiting misfit. Item 1 had several individuals who scored unexpectedly high, while Items 7 and 8 had individuals who scored unexpectedly low. Unlike the PAS pilot, there was not a single individual who appeared to be driving the misfit for these items; rather, there were several individuals throughout the sample selecting unexpected responses. No clear patterns were apparent among these respondents, so the precise cause of the misfit cannot be diagnosed.

Category Characteristic Curves

Figures 4.17 and 4.18 present the category characteristic curves (CCCs) for each of the OCAS, which show the probability that an individual will select a specific response category based upon the difference between the person and item estimates in the Rasch logit unit (see Chapter 2). These differences are shown on the *x*-axis and the response option selection probabilities are shown on the *y*-axis.

Physical Accessibility Scale.

Figure 4.17 presents the PAS CCC. The figure shows that all five response categories have the highest probability of being selected at different points along the construct continuum. Here again there is improvement from the pre-pilot. The pre-pilot results showed that few people made use of the intermediate score categories 2 and 4 (*a little more accessible than X* and *X is a little more accessible*); this no longer appears to be the case. Not only do all five response categories have an area of the continuum where they are most likely to be chosen, but the probability of selecting the most likely category in each of these areas is greater than .5.

The intersections of each response category curve represent the Andrich thresholds, which were also illustrated in the variable map presented in Figure 4.14. These thresholds indicate the logit estimate where an individual has a .5 probability of selecting one response category over another. For example, the Andrich threshold for response category 2 (*X is a little more accessible*) is the logit estimate where individuals have a .5 probability of selecting category 2 or category 1 (*X is much more accessible*). These thresholds, their fit statistics, and the average logit estimates of individuals making using of each category, are presented in Table 4.17. The Andrich thresholds have spread out considerably from the pre-pilot to the pilot study, providing evidence of increased continuum coverage by the scale items and response categories. The average estimates are also in their intended progression from response category 1 through response category 5. Fit statistics remain greater than the 1.3 criterion for response category 1; as with the pre-pilot, this may be attributable to the low frequency of use for that response category.

Figure 4.17

Physical Accessibility Scale Category Characteristic Curve (Pilot)

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
R  1.0 +                                                             +
O      |                                                             |
B      |                                                            5|
A      |111                                                       55 |
B   .8 +   1                                                    55   +
I      |    11                                                55     |
L      |      1                                              5       |
I      |       11                                           5        |
T   .6 +         1      222222                   444       5         +
Y      |          1   22      2               444   44    5          |
    .5 +           122         22  3333333   4        4455           +
O      |           21            23       334          544           |
F   .4 +         22  1          332       443         5   4          +
       |        2     1        3   2     4   33      5     44        |
R      |      22       11    33     22  4      3    5        4       |
E      |     2           1  3         24        3  5          44     |
S   .2 +   22             **         442         **             44   +
P      |222              3  1       4   22     55  33             44 |
O      |              333    111 444      22 55      33             4|
N      |          3333       444*111     555*222       33333         |
S   .0 +*********************5555555*****1111111*********************+
E      -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
       -5    -4    -3    -2    -1     0     1     2     3     4     5
        PERSON [MINUS] ITEM MEASURE
```

Table 4.17

Physical Accessibility Scale Andrich Thresholds and Average Estimates (Pilot)

Response Category	Andrich Threshold	Response Frequency	INFIT	OUTFIT	Average Estimates
1 (X is much more accessible)	N/A	18	1.96	1.78	-2.01
2 (X is a little more accessible)	-3.07	59	1.22	2.32	-1.11
3 (about the same as X)	-.76	93	.86	.80	.37
4 (a little more accessible than X)	.86	155	.98	.77	2.17
5 (much more accessible than X)	2.97	242	.76	.83	4.55

Social Engagement Scale. Figure 4.18 presents the SES CCC. The figure shows that all five response categories have the highest probability of being selected at different points along the construct continuum. Like the PAS, the SES CCC has also improved from the pre-pilot study. The area of the continuum within which a respondent is most likely to select the middle category has expanded considerably. Additionally, the most likely response option for each area of the continuum has a selection probability of at least .5.

The intersections of each response category curve represent the Andrich thresholds, which were also illustrated in the variable map presented in Figure 4.16. These thresholds, their fit statistics, and the average logit estimates of individuals making using of each category, are presented in Table 4.18. The spread between these thresholds has increased between the pre-pilot and pilot administrations, and they remain in the intended order. The threshold is lowest for the transition between categories 1 and 2, and is highest for the transition between categories 4 and 5. As with the pre-pilot, the fit statistics for response category 1 indicate some degree of misfit. However, there is one disordering among the observed average estimates; the average estimate for response category 2 (*X is a little more engaged*) is actually lower than that for response category 1 (*X is much more engaged*). This disordering, along with the high fit statistics for response category 1, may be attributable to the small number of times that response category 1 was utilized across the sample.

Figure 4.18

Social Engagement Scale Category Characteristic Curve (Pilot)

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
R  1.0 +                                                             +
O      |                                                             |
B      |11                                                          5|
A      |  11                                                      55 |
B   .8 +    11                                                  55   +
I      |      11                                               5     |
L      |        1                                            55      |
I      |         1                 33333                    5        |
T   .6 +          1              33     33                 5         +
Y      |           1            3         33     44444    5          |
    .5 +            1   222    3            3  44     44 5           +
O      |             *22   2223              *4         *4           |
F   .4 +            2 1      32             4 33       5  4          +
       |          22   1   33  22         44    3     5    44        |
R      |        22      1 3      2       4       3  55       4       |
E      |       2         *        22    4         35          44     |
S   .2 +     22         3 11        2 44          533           44   +
P      |  222         33    1       4*2         55   3            44 |
O      |22          33       111  44   222    55      333           4|
N      |        3333         444**1       ****2          3333        |
S   .0 +*********************555555*******11111**********************+
E      -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
       -5    -4    -3    -2    -1     0     1     2     3     4     5
        PERSON [MINUS] ITEM MEASURE
```

Table 4.18

Social Engagement Scale Andrich Thresholds and Average Estimates (Pilot)

Response Category	Andrich Threshold	Response Frequency	INFIT	OUTFIT	Average Estimates
1 (X is much more engaged)	N/A	9	2.77	2.77	-.78
2 (X is a little more engaged)	-2.76	29	.95	.95	-.95
3 (about the same as X)	-1.47	112	.86	.82	.51
4 (a little more engaged than X)	1.23	143	.81	1.13	2.20
5 (much more engaged than X)	3.00	183	1.02	1.03	4.30

Residual Analysis

The final piece of analysis for each of the OCAS was a principal components analysis (PCA) on the Rasch residuals. As with the pre-pilot, this procedure checks for evidence of an uncaptured construct dimension within the residuals.

Physical Accessibility Scale. Table 4.19 presents the eigenvalues and percentages of variance explained obtained from the PCA of the PAS residuals alongside eigenvalues obtained from a PCA of randomly generated data. There is encouraging improvement from the pre-pilot. Although a component with an eigenvalue greater than 2 was extracted from the PAS residuals, the difference in variance explained by this component and the first component extracted from the random data is less than 10%; this first component also has an eigenvalue and percentage of variance explained that is less than what was obtained in the pre-pilot. This finding is complemented by the scree plots for the PAS residuals and random data, which are presented in Figures 4.19 and 4.20, respectively. Although there is somewhat of a bend in the PAS residual scree plot compared to the random data, there is no visual evidence of a clear unidimensional PCA solution. Finally, the loading plot for the first two residual components again approximates a circular pattern, indicating that there is not a clear underlying component that they capture (Figure 4.21).

Table 4.19

Principal Components Analysis Results for PAS Residuals and Random Data (Pilot)

PAS Residuals			Randomly Generated Data		
Component Number	Eigenvalue	% Variance Explained	Component Number	Eigenvalue	% Variance Explained
1	2.230	24.781	1	1.572	17.466
2	1.535	17.059	2	1.278	14.198
3	1.144	12.709	3	1.226	13.621
4	1.121	12.460	4	1.086	12.063
5	.925	10.275	5	.990	11.001
6	.785	8.724	6	.896	9.958
7	.717	7.962	7	.826	9.180
8	.504	5.604	8	.661	7.343
9	.038	.427	9	.465	5.169

Figure 4.19

Scree Plot for PAS Residuals (Pilot)

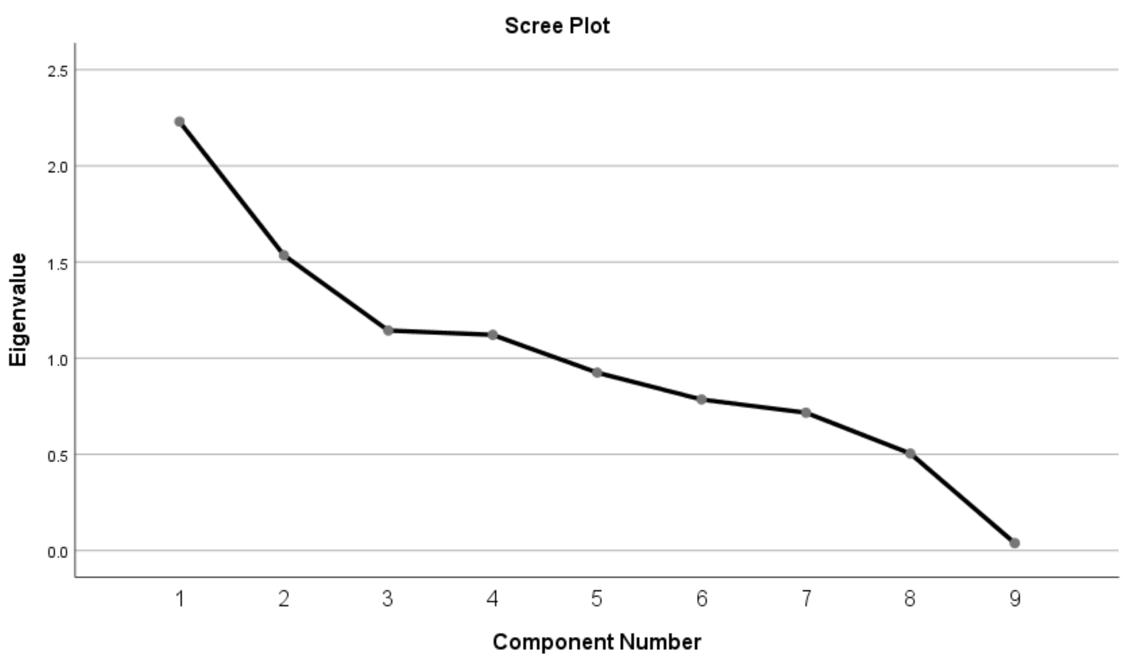


Figure 4.20

Scree Plot for Randomly Generated Data (PAS Comparison)

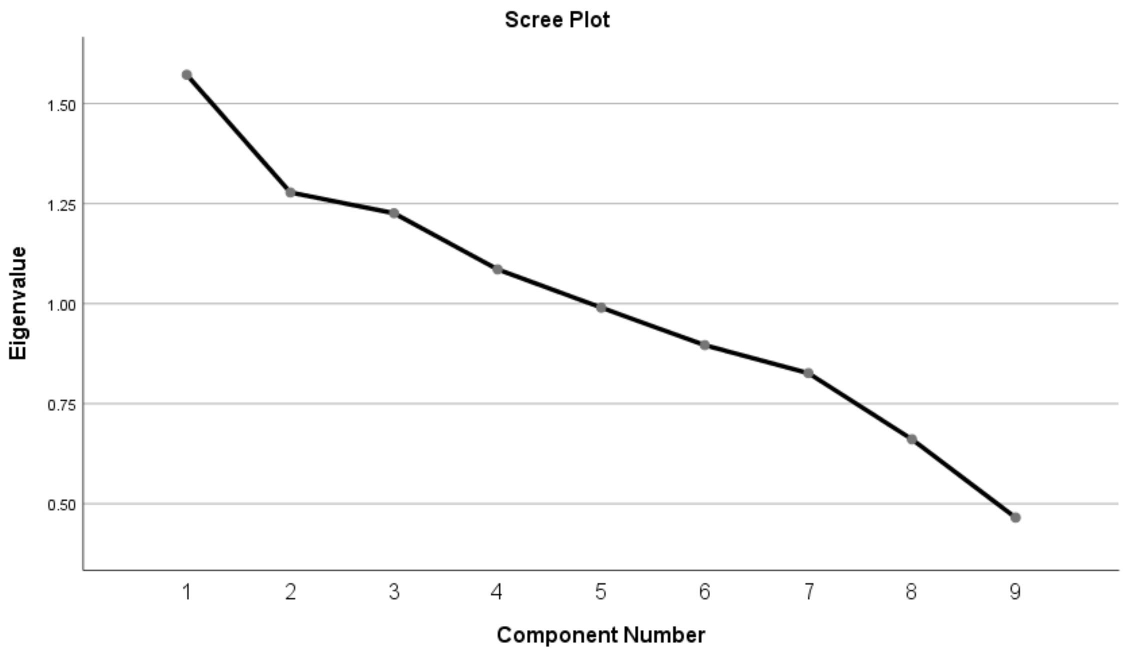
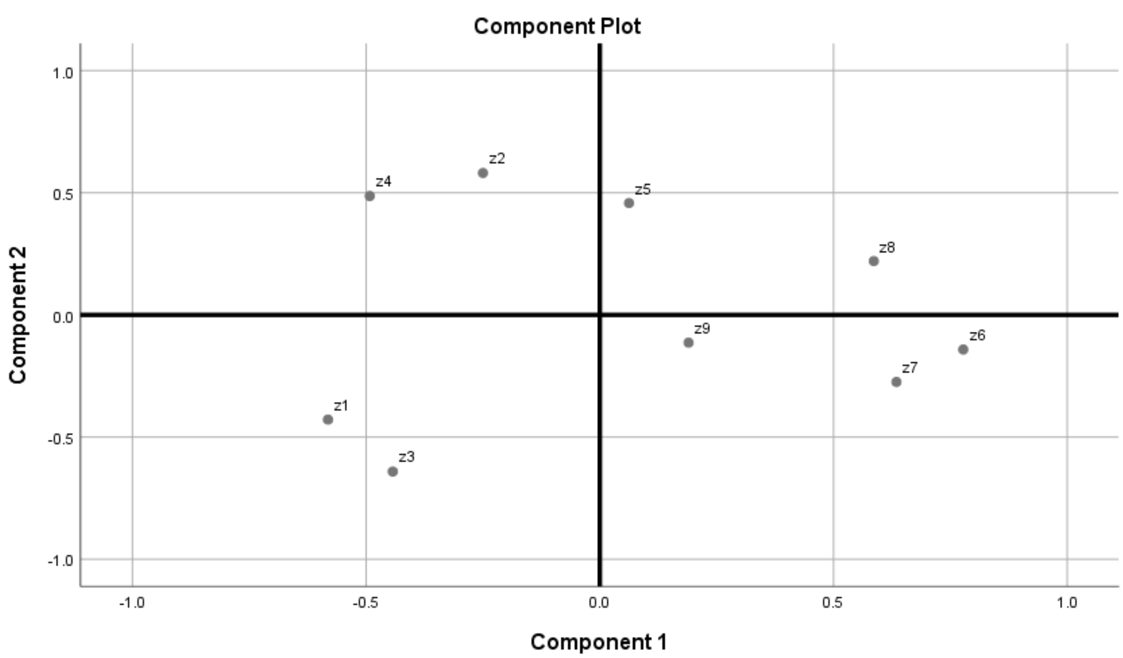


Figure 4.21

Physical Accessibility Scale Residual Component Loading Plot (Pilot)



Social Engagement Scale. Table 4.20 presents the eigenvalues and percentages of variance explained obtained from the PCA of the SES residuals alongside with eigenvalues obtained from a PCA of randomly generated data. These results have also improved from the pre-pilot. No components with eigenvalues greater than 2 were extracted from the SES residuals; additionally, the eigenvalues and percentages of variance explained by each component are similar across the residuals and the random data. This finding is also complemented by the scree plots presented in Figures 4.22 and 4.23. Finally, the loading plot of the first two residual components has also improved substantially (Figure 4.24). The items are no longer clustered in two groups and are a much better approximation of the random circular pattern.

Table 4.20

Principal Components Analysis Results for SES Residuals and Random Data (Pilot)

SES Residuals			Randomly Generated Data		
Component Number	Eigenvalue	% Variance Explained	Component Number	Eigenvalue	% Variance Explained
1	1.891	21.015	1	1.727	19.194
2	1.477	16.416	2	1.369	15.216
3	1.370	15.227	3	1.179	13.102
4	1.239	13.772	4	1.117	12.415
5	1.069	11.882	5	.979	10.875
6	.741	8.230	6	.959	10.654
7	.723	8.029	7	.642	7.132
8	.459	5.103	8	.583	6.483
9	.029	.326	9	.444	4.930

Figure 4.22

Scree Plot for SES Residuals (Pilot)

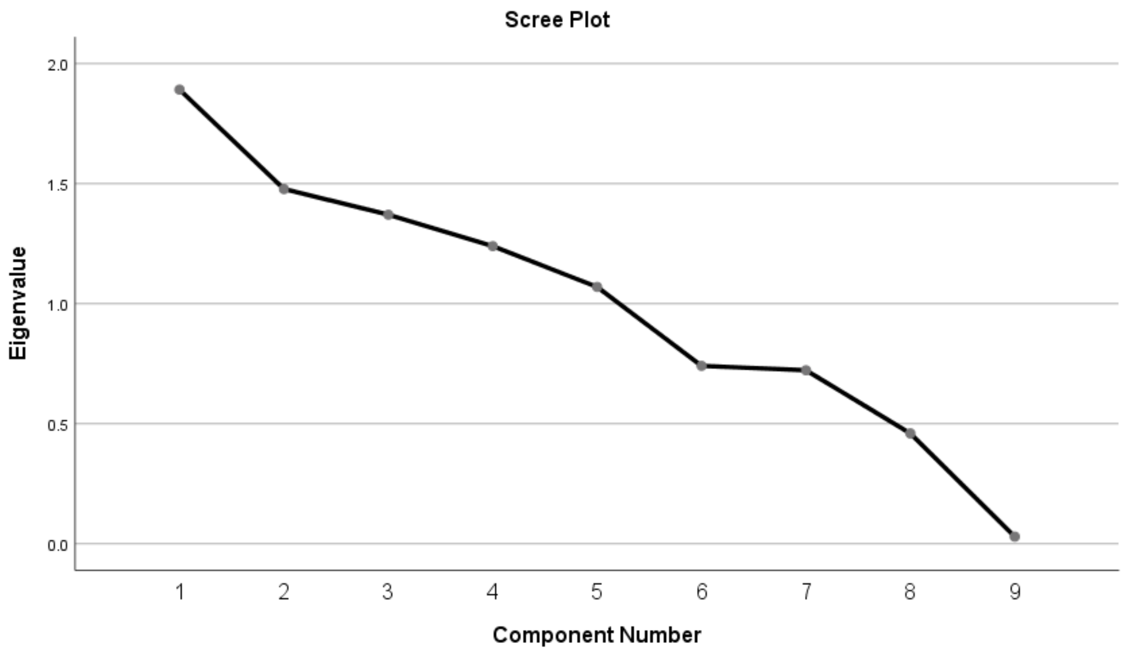


Figure 4.23

Scree Plot for Randomly Generated Data (SES Comparison)

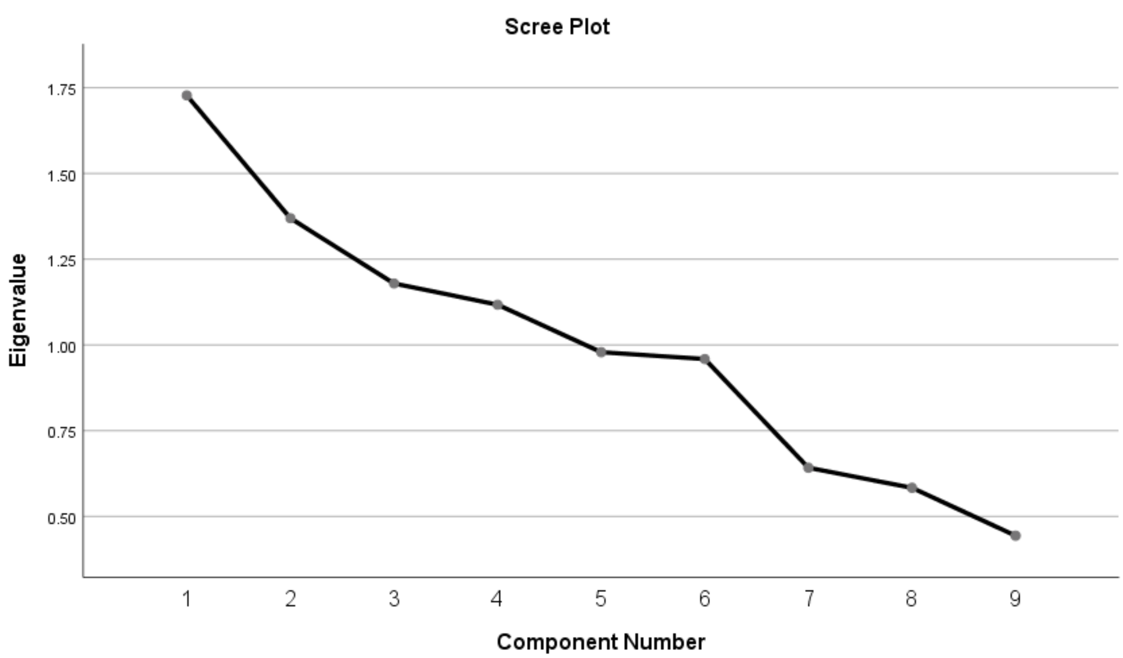
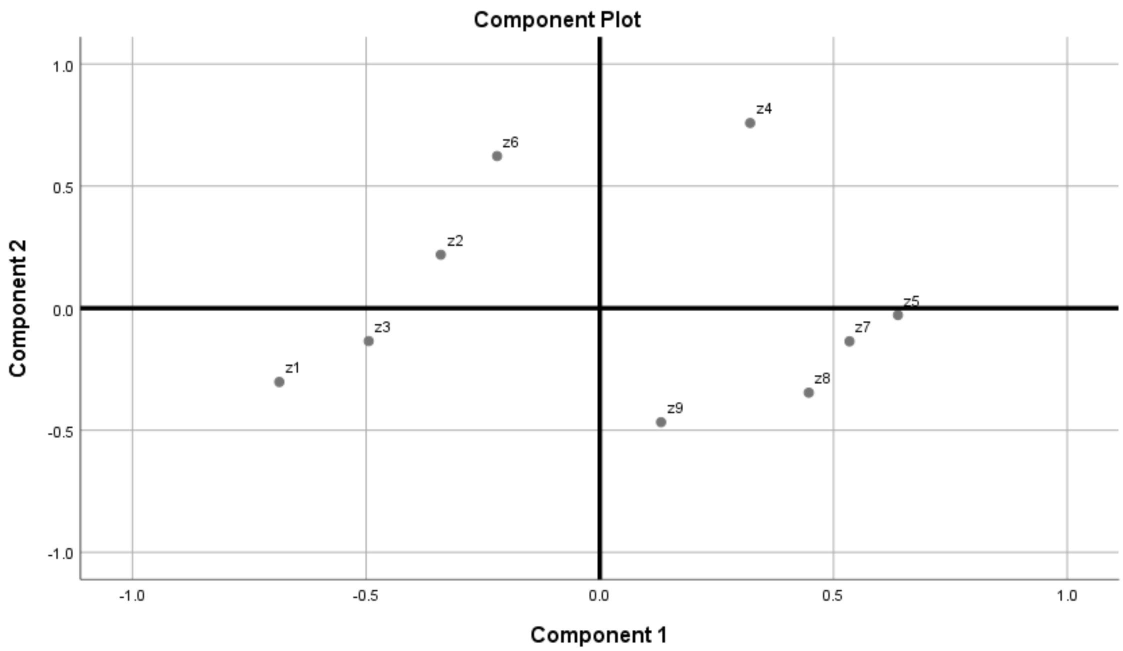


Figure 4.24

Social Engagement Scale Residual Component Loading Plot (Pilot)



Item Revisions

Additional revisions were made to scale items based upon the results of the pilot administration. Several changes were made across all items in order to address concerns about bias related to gender and perceived ethnicity of faculty names within the scenarios that were raised by a member of the dissertation committee. These changes include:

1. Removal of gendered pronouns from all scenarios. As noted in Chapter 3, scenarios were initially drafted with a balanced number of male and female faculty members for the purposes of balanced representation; however, this balance does not allay potential concerns about gender-related bias in student responses. To remedy this, rather than using "she" or "he" to refer to scenario faculty members, the gender-neutral, singular "they" pronoun was adopted. A note about this was added to survey instructions for respondents who are

unfamiliar with this usage. This adjustment required minor wording changes across all scenarios so as to clearly differentiate between when "they" was used to refer to the fictitious faculty member versus the hypothetical students within the scenario.

2. Removal of surnames for fictitious faculty members in each scenario. Similar to gender, surnames for the scenarios were initially selected to reflect a balance of different ethnicities; however, this does not address the potentiality for bias in student responses. Thus, surnames for the final scenarios were replaced with initials—for example, "Professor Connelly" became "Professor C", "Professor Atkins" became "Professor A", and so on. This construction was deemed preferable to the "Professor X" construction used in the very first drafts of the scenarios because it allows for clearer differentiation among the scenarios and allows language to flow more smoothly across all scenarios.

In addition to these changes across all items, I also made several changes to address some of the issues related to item spread from the pilot. This included dropping two items from each scale. Recall from Chapter 3 that two areas of the hypothesized continua for the PAS and SES did not have clear theoretical justifications for adjusting levels of particular facets. While this construct-related theoretical justification was lacking at the outset of scale construction, the empirical results of the pre-pilot and pilot studies provided measurement-related theoretical justification for the removal of redundant items. The items removed from each scale, along with the rationales for their removal, appear in Tables 4.21 and 4.22. These tables contain the pilot versions of each item, which do not reflect the pronoun and surname changes described above.

Table 4.21

Items Removed from Final Physical Accessibility Scale

Item	Rationale for Removal
HMH: Professor Liu holds office hours multiple days each week and meets with students via videochat if needed. She also responds quickly to emails; students know they will often receive a response within 24 hours. However, students do not usually see her around campus.	In the pre-pilot, this item and Item HHH were at the same place on the variable map, indicating that there was not much difference in responses across the two items. Although this was improved in the pilot, Item HMH and HHH were still quite close together on the variable map, again suggesting redundancy.
MLM: Professor Smith has scheduled office hours, but sometimes is not available during those times. He sometimes responds to emails within 24 hours, but other times students may wait over a week for a response. Students seldom see him around campus outside of class.	In the pre-pilot, this item was disordered, appearing above Item MMM on the variable map. This disordering remained in the pilot study.

Table 4.22

Items Removed from Final Social Engagement Scale

Item	Rationale for Removal
HMH: Professor Howard cares about meetings with students; he is attentive and does not make students feel as though they are intruding. His responses to emails address student concerns in a professional tone. However, he does not consistently acknowledge students when he sees them around campus.	In both the pre-pilot and pilot studies, items HMH and HHM were quite close together on the variable maps, suggesting some degree of redundancy.
MLM: Professor Larson is generally focused during meetings with students, although she may not always engage in thoughtful conversations. Her responses to student emails are not always clear, but are always respectful. Professor Larson rarely acknowledges students when she encounters them outside of class.	In both the pre-pilot and pilot studies, items on the lower end of the social engagement continuum were clustered closely together. Item MLM was removed to reduce redundancy and increase item spread.

Items removed from both scales were designed to reflect the same level of accessibility or engagement for the arranged meetings and email facet, with the differing level appearing in the chance encounters facet. For the PAS, these adjustments did not seem to reflect meaningful differences in students' perceptions; a faculty member being less accessible for chance encounters did not seem to affect students' perceptions in a meaningful way. These differences were not observed in the same way for the SES; however, these items were still removed for the final administration to reduce redundancy and improve data-model fit. These kinds of decisions involve tradeoffs. A greater number of scale items obviously means that more information is collected from respondents. However, given the potential burden of engaging with the scenario items, and the fact that the OCAS are two scales presented together, I erred on the side of parsimony and chose to trim items. After these removals, respondents are presented with fourteen scenarios in total: seven for each of the OCAS.

Tables 4.23 and 4.24 present the pilot versions of each scale's items juxtaposed with the final scale items.

Table 4.23

Physical Accessibility Scale Item Revisions Based on Pilot

Pilot Item	Final Item
Professor Connelly holds regular office hours and meets students outside of these times when necessary. She is a constant presence on campus and students often see her in the halls of the building where her office is located. Students can count on her to respond to emails in a timely manner.	Professor C holds regular office hours and meets students outside of these times when necessary. They seem to always be on campus and are often seen by students. Students can count on Professor C to respond to emails in a timely manner.
Professor Atkins does not hold regular weekly office hours, but often meets with students either in-person or through	Professor A does not hold regular weekly office hours, but often meets with students either in-person or through videochat.

videochat. He is on campus every day and students frequently see him outside of class, although he is inconsistent in responding to emails in a timely manner.	They are on campus every day and students frequently see Professor A outside of class. However, Professor A is inconsistent in responding to emails in a timely manner.
Professor Holfield holds office hours twice a week and will only meet with students during these times. Students only see him outside of class occasionally, but he usually responds to emails within one weekday.	Professor H holds office hours twice a week and will only meet with students during these times. Students only see them outside of class occasionally, but Professor H usually responds to emails within one weekday.
Professor Mason cannot always quickly accommodate requests she receives for meetings; students sometimes must wait over a week to meet with her. She sometimes crosses paths with students around campus and usually responds to emails in a timely manner, but students sometimes do not receive a reply for several days.	Professor M cannot always quickly accommodate meeting requests they receive; students sometimes must wait over a week. They sometimes cross paths with students around campus and usually responds to emails in a timely manner, but students sometimes do not receive a reply for several days.
Professor Olster has scheduled office hours multiple days throughout the week, but sometimes cancels them at the last minute. Students occasionally see her around campus outside of class, but she rarely responds to emails in a timely manner; students often find their emails go unanswered.	Professor O has scheduled office hours multiple days throughout the week, but sometimes cancels them at the last minute. Students occasionally see them around campus outside of class, but Professor O rarely responds to emails in a timely manner; students often find emails go unanswered.
Professor Johnson does not have scheduled office hours and typically does not accommodate student requests for meetings; students rarely see her outside of class at all. She does respond to student emails, but not always in a timely manner.	Professor J does not have scheduled office hours and typically does not accommodate student requests for meetings; students rarely see them outside of class at all. Professor J responds to student emails, but not always in a timely manner.
Professor Mullins is rarely available to meet with students; he keeps few office hours and does not accommodate requests to meet at other times. The only time students see him is during class. He rarely responds to emails in a timely manner and sometimes does not respond at all.	Professor B is rarely available to meet with students; they keep few office hours and do not accommodate requests to meet at other times. The only time students see Professor B is during class. Professor B rarely responds to emails in a timely manner and sometimes does not respond at all.

Table 4.24

Social Engagement Scale Item Revisions Based on Pilot

Pilot Item	Final Item
Professor Bradley is clearly engaged during meetings with students; he listens to students carefully, asks questions, and engaged in thoughtful discussion. He seems happy to see students around campus and often stops to chat with them. Professor Bradley's responses to emails are always courteous and clearly address students' questions or concerns.	Professor K is clearly engaged during meetings with students; they listen to students carefully, ask questions, and engage in thoughtful discussion. They seem happy to see students around campus and often stop to chat. Professor K's responses to emails are always courteous and clearly address students' questions or concerns.
Professor Jenkins clearly prioritizes students during meetings and does not seem at all distracted by other matters. She often greets students when she encounters them around campus. Her responses to emails are sometimes abrupt, but never rude or disrespectful and usually adequately address students' concerns or questions.	Professor L clearly prioritizes students during meetings and does not seem at all distracted by other matters. They often greet students they see them around campus. Professor L's responses to emails are sometimes abrupt, but never rude or disrespectful and usually adequately address students' concerns or questions.
Professor Thomas listens to students' concerns and questions during meetings, but does not always engage in thoughtful conversation. She acknowledges students inconsistently when they cross paths around campus. Professor Thomas's responses to students' emails, however, are thorough and respectful.	Professor T listens to students' concerns and questions during meetings, but does not always engage in thoughtful conversation. They acknowledge students inconsistently when crossing paths around campus. Professor T's responses to students' emails, however, are thorough and respectful.
Professor Medina occasionally seems distracted during meetings with students. He may acknowledge students when he sees them around campus, but rarely starts up a conversation. Professor Medina's responses to emails are courteous, although he does not always clearly address students' questions and concerns.	Professor Y occasionally seems distracted during meetings with students. They may acknowledge students around campus, but rarely start up a conversation. Professor Y's responses to emails are courteous, although they do not always clearly address students' questions and concerns.
Professor Simmons is not always fully engaged during meetings with students; his attention sometimes appears to be elsewhere. When he sees students around campus, Professor Simmons occasionally	Professor S is not always fully engaged during meetings with students; their attention sometimes appears to be elsewhere. When they sees students around campus, Professor S occasionally

offers a wave of recognition, but does not go out of his way to engage in conversation. His responses to emails are often unclear and may appear rude in tone.	offers a wave of recognition, but does not go out of their way to engage in conversation. Professor S's responses to emails are often unclear and may appear rude in tone.
Professor Taylor is often preoccupied with other tasks during meetings with students. She does not go out of her way to greet students around campus and seems annoyed if students try to initiate conversation. However, Professor Taylor's responses to emails are generally clear and professional.	Professor G is often preoccupied with other tasks during meetings with students. They do not go out of their way to greet students around campus and seems annoyed if students try to initiate conversation. However, Professor G's responses to emails are generally clear and professional.
Professor Roberts often seems annoyed during meetings with students and would clearly rather spend time on other activities. He does not acknowledge students when he sees them around campus and his responses to emails are often disrespectful in tone, and do not clearly address student concerns or questions.	Professor R often seems annoyed during meetings with students and would clearly rather spend time on other activities. They do not acknowledge students around campus and their responses to emails are often disrespectful in tone, and do not clearly address student concerns or questions.

Final Administration

Overview of Responses

3,000 Boston College undergraduate students were invited to complete the final administration survey. These students were randomly sampled using a list obtained from Boston College's Office of Institutional Research and Planning. The final instrument was administered using Qualtrics and data were collected over a 6-week period during October and November 2019. This administration serves to evaluate final scale properties and is also used to address my dissertation's secondary research questions.

Missing Data

Qualtrics captured 358 submitted surveys for the final administration. Of these 358 records, 121 (33.8%) were completely blank (i.e., no item responses were recorded). These 121 recorded responses were excluded at the outset of final analyses, yielding a starting sample size of 237. Of the 237 recorded responses containing data, 189 (79.7%) were complete (i.e., all seven items for both scales had recorded selections).

For the 48 cases that were not complete, 40 (16.9% of total) records contained complete data for the Physical Accessibility Scale (PAS), but no data for the Social Engagement Scale (SES). Five respondents did not complete all PAS items: two individuals responded to only a single item, two individuals responded to five items, and one individual responded to six items. Three respondents did not complete all SES items: one individual responded to three items and two individuals responded to six items. A summary of missing data from the final administration is presented in Table 4.25.

Table 4.25

Missing Data Summary (Final)

Physical Accessibility Scale			
Number of Missing Variables	Frequency (Number of Cases)	Percent	Cumulative Percent
6	2	0.8	0.8
2	1	0.4	1.2
1	2	0.8	2.0
0	232	97.9	100.0
Social Engagement Scale			
Number of Missing Variables	Frequency (Number of Cases)	Percent	Cumulative Percent
7	45	18.9	18.9
4	1	0.4	19.3
1	2	0.8	20.1
0	189	79.7	100.0

This is a small and manageable amount of missing data. The twelve respondents who completed only the PAS items most likely quit the Qualtrics survey after completing the first screen, as the PAS and SES were presented on two different pages. These individuals were retained for the PAS Rasch analysis, but excluded from the SES analysis. The two individuals who only completed a single PAS item were also excluded from all analyses because they did not provide a sufficient amount of information to adequately capture their perceptions. However, those individuals missing data for only one or two PAS items are retained for the PAS Rasch analyses because of the Rasch model's capabilities for addressing missing data. For the SES, the individual responding to only three items was excluded because they responded to fewer than half of the scale items, while the two individuals missing only a single item response were retained. After these exclusions, the two scales are left with different sample sizes for the final analyses: 235 individuals for the PAS and 191 individuals for the SES. Seven individuals selected

5 (the highest category) for all PAS items, and six individuals did so for the SES items. These respondents are automatically excluded from the Rasch analyses, but are included in the descriptive statistics.

Descriptive Statistics

Respondents

Respondents were not asked to provide demographics. This information was obtained through the records of Boston College's institutional research office. This demographic information is not associated with survey responses in order to protect students' privacy. This does limit my ability to diagnose differential responding among student groups; however separation of demographic information from item responses is a condition of my IRB approval. Table 4.26 provides information regarding respondents' gender, race/ethnicity, and the college in which they are enrolled. Race/ethnicity categories are those used in the Integrated Postsecondary Education Data System (IPEDS). Demographics are presented only for the 237 respondents who are included in the Rasch analyses. Table 4.27 presents the demographic percentages for all 3,000 survey invitees for comparison.

Table 4.26

Final Administration Respondents' Demographic Information

	Number of Respondents	Percentage of Respondents
Gender		
Male	65	27.7
Female	170	72.3
Race/Ethnicity		
Asian	42	17.9
Black or African American	6	2.6
Hispanic of Any Race	27	11.5
International	10	4.3
Race/Ethnicity Unknown	12	5.1
Two or More Races	15	6.4
White	123	52.3
College		
Arts and Sciences	153	65.1
Education	17	7.2
Management	53	22.5
Nursing	12	5.1

Table 4.27

Final Administration Invitees' Demographic Information

	% of Invitees
Gender	
Male	46.6
Female	53.4
Race/Ethnicity	
Asian	10.3
Black or African American	3.6
Hispanic of Any Race	10.7
International	8.0
Race/Ethnicity Unknown	4.2
Two or More Races	4.4
White	58.6
College	
Arts and Sciences	64.8
Education	6.2
Management	24.4
Nursing	4.6

Scenario Items

Table 4.28 presents the means and standard deviations for each of the PAS and SES scenario items. Items are presented within the table in their hypothesized order from most difficult (high level across all three facets) to least difficult (low level across all three facets). Items are denoted in the same manner as for the pre-pilot and pilot analyses: “H” indicates a high facet level, “M” indicates a medium facet level, and “L” indicates a low facet level. The number in parentheses indicates the overall “level” of the item as calculated by summing the individual facet levels (see Chapter 2 for further explanation). Items are numbered within the table for reference purposes only; items were randomized when presented to respondents.

As intended, the PAS item with the highest mean is Item 7 (LLL). This result is paralleled in the SES. As with the pre-pilot and pilot studies, the response option coding scheme is 5 = *much more accessible/engaged than X*, 4 = *a little more accessible/engaged than X*, etc. This coding scheme means that the highest item means indicate that a particular item was relatively easy for respondents to select the highest level option of *much more accessible/engaged than X*. In contrast, the items with the lowest means are those for which it was most difficult for respondents to select the highest level option of *much more accessible/engaged than X*. Thus, the finding that the items designed to reflect the lowest levels of accessibility and engagement have the highest means is theoretically appropriate.

Examination of the item means in conjunction with their designated facet levels provides a preview of the Rasch analysis results. Items with higher facet levels have lower means, meaning that it is more difficult for respondents to select the higher response categories that would indicate the faculty member they had in mind while completing the items is more accessible/engaged than the fictitious faculty member in the item scenario. Likewise, the items constructed to reflect low facet levels tend to have higher means, meaning that it is easier for respondents to select the higher categories. In fact, across both scales, only two items are not in their hypothesized difficulty order: Item 2 (HHM) and Item 3 (MMH) from the PAS.

Table 4.28

Descriptive Statistics for Scenario Items (Final Administration)

Physical Accessibility Scale		
	Mean	Standard Deviation
Item 1: HHH (9)	2.65	0.93
Item 2: HHM (8)	3.86	1.09
Item 3: MMH (7)	3.63	0.92
Item 4: MMM(6)	4.16	1.03
Item 5: MML (5)	4.34	1.01
Item 6: LLM (4)	4.60	0.94
Item 7: LLL (3)	4.61	0.89
Social Engagement Scale		
	Mean	Standard Deviation
Item 1: HHH (9)	2.69	0.91
Item 2: HHM (8)	3.14	0.86
Item 3: MMH (7)	3.95	0.84
Item 4: MMM(6)	4.25	0.88
Item 5: MML (5)	4.43	0.78
Item 6: LLM (4)	4.46	0.76
Item 7: LLL (3)	4.71	0.72

In addition to the order of item means within each scale, Table 4.28 also provides evidence of spread for scale items. For the PAS, item means range from 2.65 to 4.61 (range of 1.96); for the SES, item means range from 4.71 to 2.69 (range of 2.02). These ranges are slightly less than those obtained from the pilot study, but substantially larger than those from the pre-pilot. The highest and lowest item means and corresponding ranges from each administration are presented together in Table 4.29.

Table 4.29

Item Mean Ranges for All Scale Administrations

Physical Accessibility Scale			
	Pre-pilot (N = 43)	Pilot (N = 64)	Final (N = 235)
Highest Item Mean	4.53 (LLL)	4.79 (LLM)	4.61 (LLL)
Lowest Item Mean	3.28 (HMH)	2.49 (HHH)	2.65 (HHH)
Range	1.25	2.30	1.96
Social Engagement Scale			
	Pre-pilot (N = 41)	Pilot (N = 53)	Final (N = 191)
Highest Item Mean	4.46 (MMM)	4.83 (LLL)	4.71 (LLL)
Lowest Item Mean	3.22 (HHH)	2.66 (HHH)	2.69 (HHH)
Range	1.24	2.17	2.02

Table 4.29 provides insight concerning how the two scales have improved with each administration. In the pre-pilot, neither the PAS nor the SES had the theoretically intended items for their highest and lowest item means. The range of item means was quite small, even with the small sample size. The range of item means improved substantially between the pre-pilot and pilot administrations, suggesting the item revisions did achieve their desired effect of prompting increased variation in responses. The SES pilot administration had the theoretically intended items with the highest and lowest means; however, this was not achieved for the PAS. The final administration of both scales conforms to what is theoretically expected in terms of easiest and most difficult items, and has a much improved range over the pre-pilot administration. The Rasch analyses presented in the following section will build upon these initial findings in more detail.

Rasch Analyses

I performed a Rasch rating scale analysis (Andrich, 1978) on the final administration data for each scale. This final analysis had two purposes. The first was to address my dissertation's primary research question of whether students' perceptions of faculty availability outside of class could adequately be captured by Rasch/Guttman scenario scales (the OCAS). This question is answered through examination of the extent to which the items reflect the hypothesized continua of the accessibility and engagement constructs, as well as examination of various fit indices. Second, scores obtained from this final calibration will be used to address my dissertation's secondary research question concerning the relationship between students' perceptions of availability outside of class (as measured by the OCAS) and students' actual engagement in out-of-class communication with faculty members. Final administration Rasch analysis results are presented in the same order and fashion as for the pre-pilot and pilot studies.

Variable Maps

Physical Accessibility Scale. Figure 4.25 presents the final PAS variable map. The line in the center of the map represents the hierarchical continuum of students' perceptions of a faculty member's physical accessibility outside of class. Higher representations and perceptions of physical accessibility are represented at the top of the map and lower representations and perceptions of physical accessibility are represented at the bottom. Persons are represented differently from the pre-pilot and pilot studies on the final administration variable map because of the larger sample size. Rather than a three-digit number to represent each person, "#" is used to represent groups of three individuals, and "." represents one or two individuals. Items are denoted in the same way as the pre-pilot and pilot studies.

Figure 4.25

Physical Accessibility Scale Variable Map (Final)

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    6           .##  +
                     |
                     |
                     |
                     |
    5                +
                  .  |
                    T|
                     |
                     |
    4                +
              #####  |
                     |
                     |
           .####### S|
    3                +
       .###########  |
                     |T HHH (9)
          #########  |
                     |
    2        ######  +
           .####### M|
                     |
             ######  |S
              .####  |
    1          .###  +
                ###  |  MMH (7)
                  . S|
                 .#  |  HHM (8)
                 ##  |
    0            .#  +M
                  #  |  MMM (6)
                     |
                     |  MML (5)
                  . T|
   -1             #  +
                  .  |
                 .#  |S LLM (4)
                     |  LLL (3)
                  #  |
   -2                +
                     |
                  .  |
                     |T
                     |
   -3                +
                     |
                     |
                     |
                     |
   -4                +
                     |
                     |
                     |
                     |
   -5             .  +
               <less>|<freq>
```

This variable map reveals further improvement from the PAS pilot results. The space between items has increased in the middle portion of the continuum, likely because of the removal of redundant items. Additionally, Item LLL is now at the bottom of the variable map, indicating that it is the easiest item. However, there is one pair of disordered items: Item HHM and Item MMH. These items had been very close together in the pilot study; however, Item HHM was still slightly more difficult. In the final calibration, Item MMH has become more difficult. One possible reason for this is that the individual facets operate differently to influence students' perceptions of physical accessibility. This possibility, as well as a method for testing it empirically, will be discussed further in Chapter 5.

Another area of improvement is the relationship between the general locations of items and respondents. The representations of persons and items have considerably more overlap in the final calibration than in the pre-pilot or pilot studies, although there are still very few individuals at the low end of the continuum. As with the previous calibrations, the "M" marker on to the left of the center line on the variable map represents the average person estimate (1.94), while the "M" to the right of the line represents the average item estimate (set to 0.0). While the person average is still high compared to the item estimate, it is substantially lower than the pilot study average (2.47). Although the difference in person and item average estimates indicates potential misalignment between respondents and items, this is a positive finding from a content perspective because it shows that students do generally perceive their instructors to be physically accessible outside of class. This finding and potential implications will also be discussed further in Chapter 5.

One metric that did not improve between the pilot and final administration is the person separation. The pilot administration had a person separation of 2.10 with a reliability of .82; in the final administration, the person separation was 1.81 with a reliability of .77. This is a result of the removal of items from the scale; a smaller number of items caused the standard errors for the person estimates to increase, which negatively impacts separation. The item separation, however, has continued to improve from the pilot study because of the increase in sample size. The pilot study separation was 7.87 with a reliability of .92; in the final administration, the item separation is 11.84 with a reliability of .99. However, this improvement cannot necessarily be interpreted as additional evidence of construct continuum coverage from the pre-pilot because the sample sizes are not comparable.

Figure 4.26

Physical Accessibility Scale Variable Map with Andrich Thresholds (Final)

```
MEASURE    PERSON - MAP - ITEM - Andrich thresholds (modal categories if ordered)
                <more>|<rare> 2        3        4        5
    5            .##  +
                      |
                   .  |
                     T|                                              HHH     .5
                      |
                      |
    4                 +
               #####  |
                      |
                      |                                HHH     .4
                      |
            .####### S|
    3                 +
        .###########  |                                              MMH     .5
                      |T
                      |
           #########  |                                              HHM     .5
                      |
    2         ######  +
                     M|                  HHH     .3    MMH     .4    MMM     .5
            .#######  |
              ######  |
                      |S                               HHM     .4    MML     .5
               .####  |
    1           .###  +
                      |
                .###  |                                MMM     .4
                  .# S|                                              LLM     .5
                                                                     LLL     .5
                      |    HHH     .2                  MML     .4
                  ##  |                  MMH     .3
    0             .#  +M
                   #  |
                      |                  HHM     .3
                      |                                LLM     .4
                                                       LLL     .4
                   .  |
                     T|                  MMM     .3
   -1              #  +
                   .  |
                      |S   MMH     .2    MML     .3
                  .#  |
                      |
                   #  |    HHM     .2
   -2                 +
                      |                  LLM     .3
                                         LLL     .3
                      |
                   .  |    MMM     .2
                      |T
                      |    MML     .2
   -3                 +
                      |
                      |
                      |
                      |    LLM     .2
                      |    LLL     .2
   -4              .  +
                <less>|<freq> 1        2        3        4
```

Figure 4.26 shows a variable map that includes the Andrich thresholds, which represent the point at which an individual has a 50% probability of selecting the next higher response category for the item. In this map, each item appears four times: once for each response category transition. The item labels followed by ".2" are the Andrich thresholds for moving between response categories one and two (*X is much more accessible* and *X is a little more accessible*), labels followed by ".3" are the Andrich thresholds for moving between response categories two and three (*X is a little more accessible* and *about the same as X*), and so on. Persons are represented the same way as in Figure 4.25.

Figure 4.26 illustrates how many individuals in the sample were relatively high-scoring on the PAS. Most respondents were generally expected to select at least *about the same as X* (score of 3) for all items. However, the spread of response category transitions has improved between the pilot study and final administration.

Social Engagement Scale. Figure 4.27 presents the final SES variable map. The line in the center of the map represents the hierarchical continuum of students' perceptions of a faculty member's social engagement outside of class. The layout of this variable is comparable to the PAS: higher representations and perceptions of social engagement are represented at the top of the map and lower representations and perceptions of social engagement are represented at the bottom. Persons and items are denoted the same way as in the PAS.

Figure 4.27

Social Engagement Scale Variable Map (Final)

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    7            ##  +
                     |
                     |
                     |
    6                +
                     |
                     |
                 .# T|
    5                +
                     |
                 ##  |
                     |
    4                +
          .######## S|T
                     |
           .#######  |  HHH (9)
    3                +
           .#######  |
         ##########  |
                    M|  HHM (8)
    2         #####  +
              .####  |S
                     |
                 .#  |
    1         .####  +
                .## S|
                     |
                .##  |  MMH (7)
    0            .#  +M
                  .  |
                     |
                  . T|  MMM (6)
   -1             .  +
                     |  LLM (4)   MML (5)
                     |
                     |S
   -2                +
                  .  |  LLL (3)
                     |
                     |
   -3                +
                     |
                     |
                     |T
   -4             .  +
                     |
                     |
                     |
   -5             .  +
               <less>|<freq>
```

Figure 4.27 shows that the spread of SES items has improved between the pilot study and final administration, likely due to the removal of redundant items. Items MML and LLM appear next to each other on the variable map; however, inspection of item difficulty estimates reveals that they are still in their intended difficulty order and that this display is a result of WINSTEPS formatting rather than disordered items. Still, these two items are quite close together and increased spread between them would be desirable in future iterations of the scale. Another area for improvement is the relatively large gap on the variable map between items HHM and MMH. A large gap in this area means that a qualitative description of individuals' perceptions of social engagement is lacking in this area of the continuum, which compromises the overall interpretability of the scale. As with the disordered items in the PAS discussed above, these issues may be indicative of differential operation of facets within the social engagement construct. This will be discussed further in Chapter 5.

Similar to the pilot study, final administration respondents were quite high-scoring overall. The average person estimate, which represents the student's perception of their focal faculty member, was 2.48, which is substantially higher than the average item estimate that is set to 0. The mismatch between persons and items is more extreme for the SES than it was in the PAS and shows that the final administration sample generally found the faculty members they considered to be socially engaged during out-of-class interaction. From a content perspective, this is a positive finding that will be discussed further in Chapter 5.

The person separation also decreased for the SES between the pilot study and final administration. Again, this is due to the reduction in the number of scale items, which increased the standard errors for the person estimates. The pilot study person separation was 2.00 with a reliability of .80; in the final administration, this has decreased modestly to a separation of 1.89 with a reliability of .78. Unsurprisingly, the item separation has continued to increase. The pilot study item separation was 6.18 with a reliability of .97; this has increased to 13.53 and .99, respectively. This is a function of increased sample size and cannot necessarily be interpreted as evidence for coverage of the social engagement construct continuum.

Figure 4.28

Social Engagement Scale with Andrich Thresholds (Final)

```
MEASURE    PERSON - MAP - ITEM - Andrich thresholds (modal categories if ordered)
               <more>|<rare> 2        3        4        5
    7            ##  +
                     |
                     |
                     |                                              HHH     .5
    6                +
                     |
                     |
                 .# T|                                              HHM     .5
    5                +
                     |                                HHH     .4
                 ##  |
                     |
    4                +
          .######## S|T
                     |                                HHM     .4
           .#######  |                                              MMH     .5
    3                +
           .#######  |
         ##########  |                                              MMM     .5
                    M|
    2         #####  +                                              MML     .5
              .####  |S                 HHH     .3                  LLM     .5
                     |                                MMH     .4
                 .#  |
    1         .####  +
                .## S|                  HHM     .3    MMM     .4    LLL     .5
                     |
                .##  |                                MML     .4
    0            .#  +M   HHH     .2                  LLM     .4
                  .  |
                     |
                  . T|
   -1             .  +    HHM     .2                  LLL     .4
                     |                  MMH     .3
                     |
                     |S
   -2                +                  MMM     .3
                  .  |
                     |
                     |                  MML     .3
                                        LLM     .3
   -3                +    MMH     .2
                     |
                     |
                     |T   MMM     .2    LLL.3
   -4             .  +
                     |    MML     .2
                     |    LLM     .2
                     |
   -5             .  +
                     |
                     |    LLL     .2
                     |
   -6                +
               <less>|<freq> 1        2        3        4
```

Figure 4.28 shows a variable map variation that includes the Andrich thresholds, which represent the point at which an individual has a 50% probability of selecting the next higher response category for the item. The schematic of this map is comparable to Figure 4.18, which illustrates these results for the PAS. This iteration of the SES variable map again illustrates the high scores of respondents; most individuals were expected to select at least a 3 (*about the same as X*) for all scale items.

Fit Statistics

Rasch model fit statistics for the final administration were used to provide further confirmation of data/model fit. While the variable maps provide a visual representation of how the items and persons fall along the hypothesized continua, fit statistics are useful for providing more nuanced diagnostic information. As with the pre-pilot and pilot studies, only item fit statistics are discussed in detail here, as the purpose of this study is to establish the general utility of the OCAS instruments and not to provide precise estimates for individual students' perceptions.

Physical Accessibility Scale. Table 4.30 presents the fit statistics for the final PAS data. The items are listed in difficulty order, as indicated in the logit estimate column. A higher logit estimate indicates an item for which it is more difficult for respondents to select the upper-level categories (i.e., to indicate that the faculty member they have in mind is at least as accessible as the fictitious faculty member in the item). The order of items by Rasch difficulty measure is identical to the order of items by their mean responses.

Table 4.30

Physical Accessibility Scale Fit Statistics (Final)

Item	Logit Estimate (S.E.)	Information-Weighted Fit Statistic		Unweighted Fit Statistic	
		MNSQ	ZSTD	MNSQ	ZSTD
Item 1: HHH (9)	2.53 (.09)	1.02	.25	2.10	8.70
Item 3: MMH (7)	.86 (.09)	.96	-.47	1.11	1.18
Item 2: HHM (8)	.43 (.09)	1.30	3.02	1.23	2.24
Item 4: MMM (6)	-.20 (.10)	.75	-2.72	.68	-3.18
Item 5: MML (5)	-.63 (.11)	.92	-.75	.72	-2.27
Item 6: LLM (4)	-1.47 (.13)	1.28	2.04	.69	-1.76
Item 7: LLL (3)	-1.52 (.13)	1.03	.23	.59	-2.43

Six of the seven scale items have acceptable INFIT and OUTFIT statistics. Item 1, which had an OUTFIT statistic greater than 1.3, and Item 2, which had an INFIT statistic of exactly 1.3 are highlighted in the table. It is also worth noting that Item 2 is a member of the one disordered item pair that did not fall where expected along the construct continuum. During the scale development process, consistent inconsistencies in responses (captured by the INFIT statistic) are of greater concern than individual surprising responses (captured by the OUTFIT statistic); however, both of these can be investigated further by examining individual person response patterns (presented in Appendix K).

Examination of the person-response table revealed several individuals who scored unexpectedly high on Item 1, causing its high OUTFIT statistic. This finding might be interpreted as a start-up effect had the items not been randomized for presentation to respondents. Several individuals in the person-response table had high standardized residuals for both Item 1 and Item 3, which may be indicative of reversing the order of response options. However, given that Item 2 did not display the same consistent pattern, I did not feel comfortable making that extrapolation and removing these individuals from the analysis, as I did for one respondent in the pilot administration.

The response patterns also revealed inconsistency in responses for Item 2, which was expected because of the INFIT statistic. It is difficult to pinpoint a cause for this; it is possible that the phrasing of this item should be revised. However, it is also possible that this statistic is simply a function of this particular sample, as this item did not display problematic fit statistics in the pilot administration.

Social Engagement Scale. Table 4.31 presents the fit statistics for the final SES data. The items are listed in difficulty order, as indicated in the logit estimate column. A higher logit estimate indicates an item for which it is more difficult for respondents to select the upper-level categories (i.e., to indicate that the faculty member they have in mind is at least as accessible as the fictitious faculty member in the item). As with the PAS and previous administrations, the order of items by Rasch difficulty measure is identical to the order of items by their mean responses.

Table 4.31

Social Engagement Scale Fit Statistics (Final)

Item	Logit Estimate (S.E.)	Information-Weighted Fit Statistic		Unweighted Fit Statistic	
		MNSQ	ZSTD	MNSQ	ZSTD
Item 1: HHH (9)	3.24 (.11)	1.10	.96	2.19	6.73
Item 2: HHM (8)	2.14 (.11)	1.01	.10	1.36	2.84
Item 2: MMH (7)	.15 (.12)	.78	-2.45	.76	-2.45
Item 4: MMM (6)	-.64 (.12)	.90	-.97	.82	-1.54
Item 5: MML (5)	-1.23 (.13)	1.00	.03	1.17	1.14
Item 6: LLM (4)	-1.32 (.14)	1.05	.50	.87	-.84
Item 7: LLL (3)	-2.35 (.17)	1.37	2.39	.79	-.83

Six of the seven scale items have acceptable INFIT statistics, and five of the seven items have acceptable OUTFIT statistics. As noted above, individual surprising responses (captured by OUTFIT) are of less concern at the scale development stage compared to consistent inconsistencies (captured by INFIT). Items HHH and HHM, the most difficult items, both had OUTFIT statistics greater than the criterion of 1.3, suggesting that there were some surprisingly high individual responses to these scale items. Item LLL had an INFIT statistic greater than 1.3, suggesting general inconsistencies in responses. Both of these speculations can be verified through examination of the person-response tables (presented in Appendix K).

As expected, there was a small number of individuals who had unexpectedly high responses for the most difficult SES items. Because the items were presented in a randomized order, this cannot be attributed to a start-up effect. For Item LLL, several students selected responses that were lower than expected. Because this item was intended to be very easy (and most respondents selected one of the top two response options), these individuals triggered the INFIT statistic for this item. The reasons for this inconsistency are not clear upon examination of the response patterns; there is no clear evidence that students reversed the scale options or made other systematic errors.

Category Characteristic Curves

Figures 4.29 and 4.30 present the category characteristic curves (CCCs) for each of the OCAS, which show the probability that an individual will select a specific response category based upon the difference between the person and item estimates in the Rasch logit unit (see Chapter 2). These differences are shown on the *x*-axis and the response option selection probabilities are shown on the *y*-axis. These curves are evaluated to see if respondents are making use of all response categories for the scale items.

Physical Accessibility Scale.

Figure 4.29 presents the PAS CCC. The figure shows that all five response categories have the highest probability of being selected at different points along the construct continuum. These results are comparable to what was observed in the pilot data: each response option as a clear area of the construct continuum where it is most likely to be chosen. This provides evidence that the PAS response options are appropriate for capturing the ways in which respondents' perceptions may vary relative the descriptions provided in the scenario items.

The intersections of each response category curve represent the Andrich thresholds, which were also illustrated in the variable map presented in Figure 4.26. These thresholds indicate the logit estimate where an individual has a .5 probability of selecting one response category over another. For example, the Andrich threshold for response category 2 (*X is a little more accessible*) is the logit estimate where individuals have a .5 probability of selecting category 2 or category 1 (*X is much more accessible*). These thresholds, their fit statistics, and the average logit estimates of individuals making use of each category, are presented in Table 4.32. The spread of Andrich thresholds has actually decreased between the pilot study and the final administration; however, the thresholds and average estimates are in their intended order. As with the two previous administrations, the INFIT and OUTFIT statistic remain above the 1.3 criterion for response category 1. As in the pre-pilot and pilot studies, this may be because relatively few times this response option is utilized across the PAS items. Within this small frequency, individuals who were otherwise high-scoring unexpectedly chose response category 1, contributing to its misfit. These individuals can be observed in the person-response tables in Appendix K.

Figure 4.29

Physical Accessibility Scale Category Characteristic Curve (Final)

```
     CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+------+------+------+------+------+------+------+------+-
R  1.0 +                                                         +
O      |                                                         |
B      |                                                        5|
A      |11                                                    55 |
B   .8 +  11                                                55   +
I      |    11                                            55     |
L      |      1                                         55       |
I      |       11                                      5         |
T   .6 +         1                                    5          +
Y      |          1                                  5           |
    .5 +           1  2222222     333333            5            +
O      |            *2       22 33      33   4444455             |
F   .4 +          22 11       3*2         **4    5444            +
       |        22     1     3   2       4  3   5    44          |
R      |      22        1   3     22   44    3 5       44        |
E      |    22           1*3        2 4       *3         44      |
S   .2 +  22            33 1        4*2     55  3          44    +
P      |22            33    11    44   22  5     33          444 |
O      |            33        1*44      5**        333          4|
N      |       33333       4444 1111*555   2222       33333      |
S   .0 +*******************555555555 1111111111******************+
E      -+------+------+------+------+------+------+------+------+-
       -4     -3     -2     -1      0      1      2      3      4
        PERSON [MINUS] ITEM MEASURE
```

Table 4.32

Physical Accessibility Scale Andrich Thresholds and Average Estimates (Final)

Response Category	Andrich Threshold	Response Frequency	INFIT	OUTFIT	Average Estimates
1 (X is much more accessible)	N/A	67	1.74	1.67	-1.41
2 (X is a little more accessible)	-2.23	147	1.16	1.18	-.56
3 (about the same as X)	-.69	287	.79	.96	.37
4 (a little more accessible than X)	.94	393	1.00	.83	1.73
5 (much more accessible than X)	1.97	747	.97	1.06	3.31

Social Engagement Scale. Figure 4.30 presents the SES CCC. The figure shows that all five response categories have the highest probability of being selected at different points along the construct continuum. These results are comparable to what was observed in the pilot data: each response option as a clear area of the construct continuum where it is most likely to be chosen. This provides evidence that the SES response options are appropriate for capturing the ways in which respondents' perceptions may vary relative to the descriptions provided in the scenario items.

The intersections of each response category curve represent the Andrich thresholds, which were also illustrated in the variable map presented in Figure 4.28. These thresholds indicate the logit estimate where an individual has a .5 probability of selecting one response category over another. These thresholds, their fit statistics, and the average logit estimates of individuals making using of each category, are presented in Table 4.33. The spread of the Andrich thresholds has improved between the pilot study and final administration. However, response categories 1 and 2 both have INFIT and OUTFIT statistics greater than 1.3. For response category 1, these may be attributable to a low frequency of category usage across the SES items. Similar to the PAS, within this relatively small group of responses, some individuals who were otherwise high-scoring unexpectedly selected response category 1, triggering misfit. These individuals can be observed in the person-response tables in Appendix K. The reason for misfit in response category 2 is not clear.

Figure 4.30

Social Engagement Scale Category Characteristic Curve (Final)

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
R  1.0 +                                                             +
O      |                                                             |
B      |                                                             |
A      |11                                                         55|
B   .8 +  11                                                     55  +
I      |    1                                                  55    |
L      |     11                                               5      |
I      |       1                   3333333                   5       |
T   .6 +        1                33       33                5        +
Y      |         1    22222     3           3      444     5         |
    .5 +          1 22     22  3             3  444   444 5          +
O      |          2*         2*               **         *           |
F   .4 +         2  1       33 2             4  3      55 44         +
       |       22    1     3    22          4    3    5     4        |
R      |      2       11  3       2       44      3  5       44      |
E      |    22          13         2     4         35          44    |
S   .2 +  22           331          22 44          533           4   +
P      |22            3   11         4*2         55   3           444|
O      |           333      11     44   222    55      333           |
N      |       3333           ****4        ****           3333       |
S   .0 +**********************5555*********1111**********************+
E      -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
       -5    -4    -3    -2    -1     0     1     2     3     4     5
        PERSON [MINUS] ITEM MEASURE
```

Table 4.33

Social Engagement Scale Andrich Thresholds and Average Estimates (Final)

Response Category	Andrich Threshold	Response Frequency	INFIT	OUTFIT	Average Estimates
1 (X is much more engaged)	N/A	34	1.44	1.44	-2.79
2 (X is a little more engaged)	-3.15	102	1.37	1.79	-1.03
3 (about the same as X)	-1.40	304	.79	1.41	.45
4 (a little more engaged than X)	1.43	357	.84	.73	2.52
5 (much more engaged than X)	3.12	538	1.09	1.10	4.33

Residual Analysis

The final piece of analysis for each of the OCAS was a principal components analysis (PCA) on the Rasch analysis residuals. This procedure checks for evidence of an uncaptured construct dimension within the residuals.

Physical Accessibility Scale. Table 4.34 presents the eigenvalues and percentages of variance explained obtained from the PCA of the PAS residuals alongside the eigenvalues obtained from a PCA of randomly generated data. These results are similar to what was obtained in the pilot study. One component with an eigenvalue greater than 2 was extracted from the PAS residuals, it is not a clear dominant factor according to various criteria. For example, it does not explain twice the variance of the second extracted component. Additionally, four of the seven extracted components explain more than 10% of variance in the residuals. Comparing the scree plots of PAS residual components and randomly generated data also reveals that while the residuals are not completely random, there is not sufficient evidence for a single dominant component. Finally, the component loading plot reveals that the residuals approximate a random circular pattern. These are positive findings because they provide evidence that there is not an uncaptured dimension of the perceived physical accessibility construct hidden within the final PAS residuals.

Table 4.34

Principal Components Analysis Results for PAS Residuals and Random Data (Final)

PAS Residuals			Randomly Generated Data		
Component Number	Eigenvalue	% Variance Explained	Component Number	Eigenvalue	% Variance Explained
1	2.439	34.845	1	1.273	18.187
2	1.490	21.288	2	1.147	16.381
3	.998	14.255	3	1.034	14.766
4	.910	12.999	4	.992	14.172
5	.558	7.964	5	.899	12.39
6	.510	7.286	6	.883	12.610
7	.095	1.362	7	.773	11.046

Figure 4.31

Scree Plot for PAS Residuals (Final)

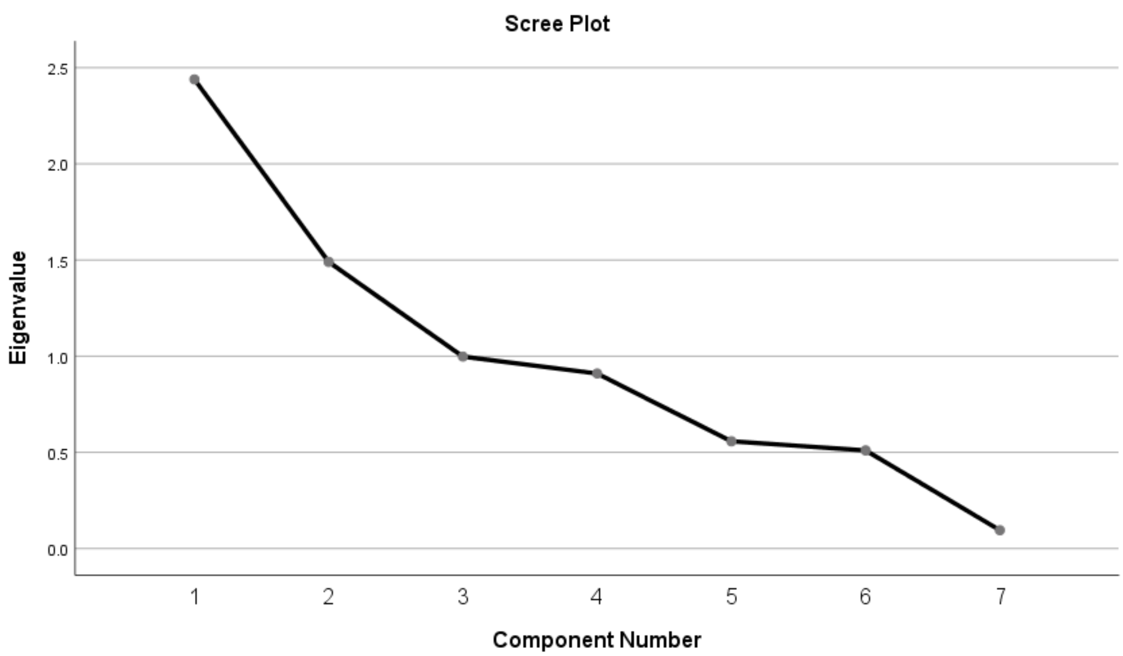


Figure 4.32

Scree Plot for Randomly Generated Data (PAS Comparison)

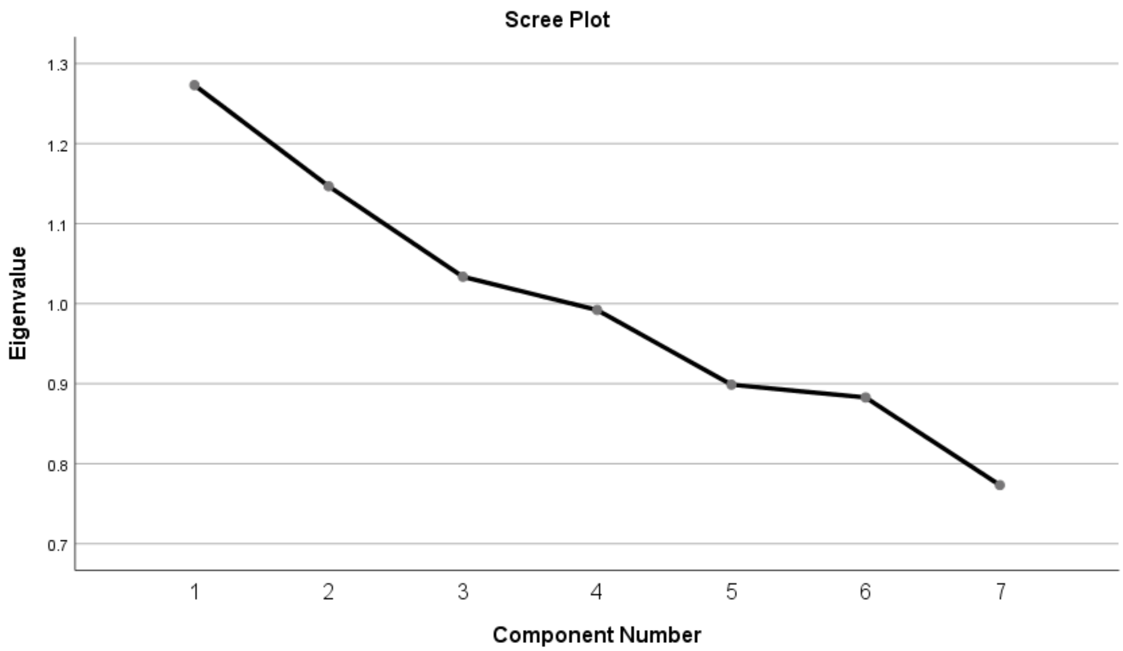
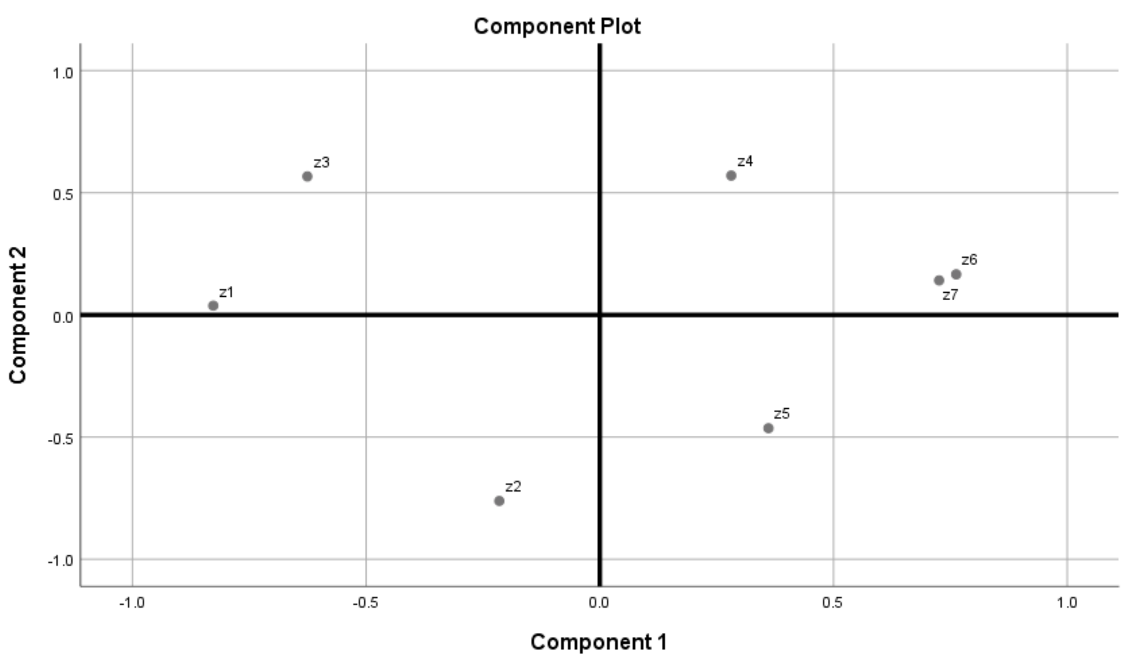


Figure 4.33

Physical Accessibility Scale Residual Component Loading Plot (Final)



Social Engagement Scale. Table 4.35 presents the eigenvalues and percentages of variance explained obtained from the PCA of the SES residuals with eigenvalues obtained from a PCA of randomly generated data. These results are slightly less random than in the pilot study; the first component extracted has an eigenvalue greater than 2. However, there is still not sufficient evidence for a strong unidimensional solution. For example, four eigenvalues of extracted components are greater than 1, and five of the seven components explain more than 10% of variance in the residuals. The residual component loading plot does roughly approximate a circular pattern; however, there are clear clusters of items within the overall pattern. Compared to the random data, the SES residual scree plot does show a bit of a bend; however, this bend does not occur near the x-axis of the plot. These findings provide evidence that there is not an uncaptured dimension of perceived social engagement in the final SES residuals, although in the future it would be beneficial to perform this analysis again with a different sample for verification.

Table 4.35

Principal Components Analysis Results for SES Residuals and Random Data (Final)

SES Residuals			Randomly Generated Data		
Component Number	Eigenvalue	% Variance Explained	Component Number	Eigenvalue	% Variance Explained
1	2.131	30.436	1	1.332	19.023
2	1.155	16.497	2	1.149	16.421
3	1.118	15.973	3	1.036	14.798
4	1.050	15.003	4	.973	13.894
5	.748	10.638	5	.919	13.124
6	.637	9.093	6	.852	12.168
7	.162	2.316	7	.739	10.563

Figure 4.34

Scree Plot for SES Residuals (Final)

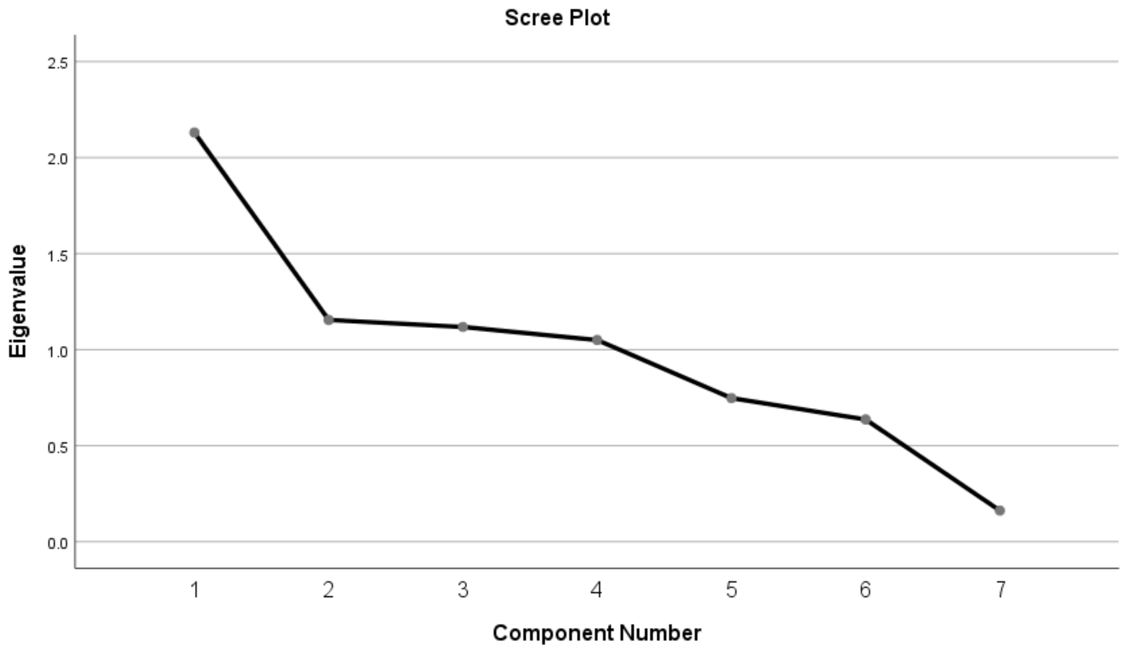


Figure 4.35

Scree Plot for Randomly Generated Data (SES Comparison)

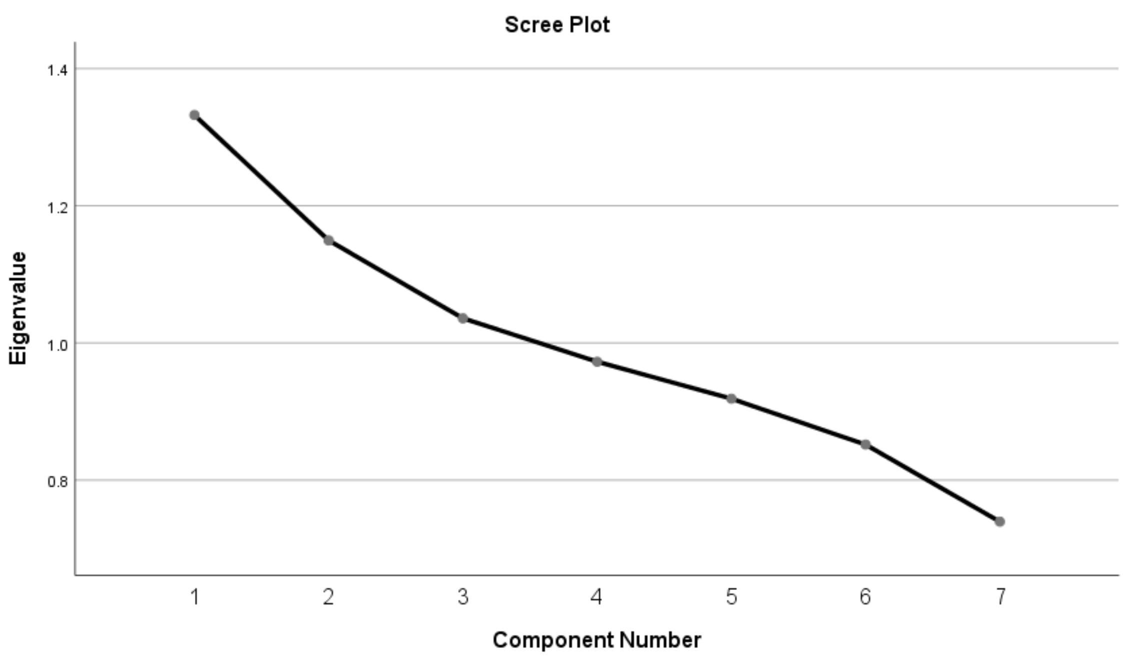
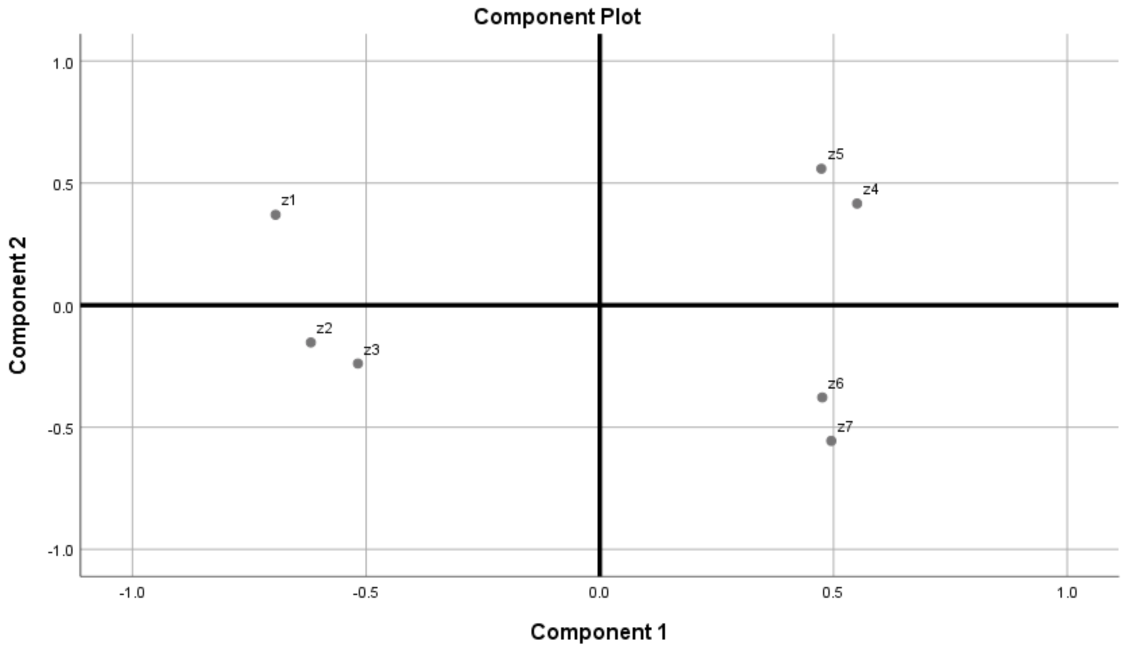


Figure 4.36

Social Engagement Scale Residual Component Loading Plot (Final)



Final Variable Maps and Interpretation

This section provides variable maps formatted to maximize the interpretability of the OCAS. As was noted in earlier chapters, the key advantage afforded by usage of the Rasch/Guttman scenario scale methodology is the interpretability of scale scores. Because the Rasch scaling procedure places items and persons along a single continuum, items may be interpreted as descriptions of the individuals who are near them on the variable map. The quality of these descriptions is enhanced for RGS scales because of the deliberate construction of scenarios as items.

Several formatting changes have been made to the variable maps presented below in order make this interpretability more apparent. The numbers on the left side of the variable maps have been converted out of the Rasch logit unit and into raw scores. The conversion tables used for the procedure are presented in Appendix L. These raw scores are simply the sum of an individual's item responses. Lines denoting where average item scores fall along the continuum have also been added to these variable maps to create raw score "zones". Items falling within each zone may be interpreted as representations of students' perceptions regarding the faculty member they had in mind while completing the scales. This is illustrated through the overlay of selected item text on the variable maps. Both of these maps utilize data from the final OCAS administration.

Physical Accessibility Scale

Figure 4.37 presents the PAS variable map with added interpretational information. The raw scale scores along the left side of the map range from 7 (indicating a selection of "much less accessible than X" for all items) to 35 (indicating a selection of "much more available than X" for all items). These raw scores allow for the division of the PAS continuum into different score "zones", for which the scenario items can serve as descriptions. These descriptions are illustrated in both Figure 4.37 and Table 4.36.

Beginning at the low end of the physical accessibility continuum, students scoring from 7-14 were expected to find the faculty member that they rated *much less accessible* than what is depicted in all seven scenarios. Individuals in this range found their focal faculty member *much less accessible* than someone who: (a) is rarely available to meet with students, (b) is only seen during class, and (c) rarely responds to emails in a timely manner and sometimes does not respond at all. Very few respondents fall in this score zone.

Moving up the continuum, individuals scoring in the 15-21 range are expected to select a *little less accessible than X* for the items in their zone and *about the same as X* for items in the lowest zone. The relatively small number of students falling in this score range reported perceiving their focal faculty members as *a little less accessible* than a faculty member who (a) cannot always accommodate meeting requests and sometimes keeps students waiting for over a week, (b) sometimes crosses paths with students around campus, and (c) usually responds to emails in a timely manner, but sometimes does not respond for several days.

Further up the PAS continuum, individuals scoring in the 22-28 range are predicted to select *about the same as X* for the items falling in their zone. These students perceived the faculty member they had in mind as about the same as an individual who (a) holds regular office hours, but will not meet students outside of them, (b) is only seen outside of class occasionally; and (c) responds to emails within one weekday. They are also predicted to select *a little more accessible than X* for items in lower zones.

Finally, respondents in the highest score zone (range of 29 to 35) were expected to select *a little more accessible than X* for the item in that zone. This item, HHH, represents high perceptions of physical accessibility across all three facets. Students in this score range are thus predicted to perceive the faculty member they had in mind while completing the PAS as *a little more accessible* than what is described by the item, and as *much more accessible* than the scenario items in lower zones. This means that they perceived the faculty member they had in mind as a little more accessible than someone who: (a) holds regular office hours and meets outside of them when necessary (b) seems to always be on campus and is often seen by students, and (c) always responds to emails in a timely manner. As has been noted in previous sections, individuals in the sample were relatively high-scoring; thus, much of the sample falls in this score range. Given the description of the individual in Item HHH, it is somewhat difficult to imagine what a faculty member might do in real life to be perceived as more accessible than this description. Exploration of this topic is an area for future research discussed in Chapter 5.

Figure 4.37

Interpretational Physical Accessibility Scale Variable Map

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
  35             .##  +   ---------------------------------------->  Average Item Score = 5
                      |
                      |
                      |
                      |
                      +
                    . |
 34                  T|
                      |
                      |
                      +
 33            #####  |
                      |
                      |
          .####### S|
                      +           HHH: Professor C holds regular office hours and
 31   .###########    |           meets students outside of these times when
                      |T HHH      necessary. They seem to always be on campus and are
           #########  |           often seen by students. Students can count on
                      |           Professor C to respond to emails in a timely
 29           ######  +           manner.
          .####### M|
 28                   |  ---------------------------------------->  Average Item Score = 4
              ######  |S
               .####  |           MMH: Professor H holds office hours twice a week
                .###  +           and will only meet with students during these
 25              ###  |  MMH      times. Students only see them outside of class
                   . S|           occasionally, but Professor H usually responds to
                  .#  |  HHM      emails within one weekday.
                  ##  |
 21               .#  +M  ---------------------------------------->  Average Item Score = 3
                   #  |  MMM
                      |           MMM: Professor M cannot always quickly accommodate
                      |  MML      meeting requests they receive; students sometimes
                   . T|           must wait over a week. They sometimes cross paths
 17                #  +           with students around campus and usually responds to
                   .  |           emails in a timely manner, but students sometimes
                  .#  |S LLM      do not receive a reply for several days.
 14                   |  LLL  ------------------------------------>  Average Item Score = 2
                   #  |
 13                   +
                      |           LLL: Professor B is rarely available to meet with
                   .  |           students; they keep few office hours and do not
                      |T          accommodate requests to meet at other times. The
                      |           only time students see Professor B is during class.
                      +           Professor B rarely responds to emails in a timely
 10                   |           manner and sometimes does not respond at all.
                      |
                      |
  9                   |
                      +
                      |
                      |
                      |
  8                   |
                   .  +
  7            <less>|<freq>  ----------------------------------->  Average Item Score = 1
```

Table 4.36

Physical Accessibility Scale Score Interpretations

Score or Range	Perception Level	Description	Illustrative Scenario
35	Extremely accessible	Student rates faculty member they had in mind as much more accessible than all scenarios.	N/A
29-34	Very accessible	Student rates faculty member as a little more accessible than HHH and much more accessible than lower scenarios.	Professor C holds regular office hours and meets students outside of these times when necessary. They seem to always be on campus and are often seen by students. Students can count on Professor C to respond to emails in a timely manner.
22-28	Moderately accessible	Student rates faculty member as about the same as HHM or MMH and as a little more accessible than lower scenarios.	Professor H holds office hours twice a week and will only meet with students during these times. Students only see them outside of class occasionally, but Professor H usually responds to emails within one weekday.
15-21	Moderately inaccessible	Student rates faculty member as a little less accessible than MMM or MML and much less accessible than lower scenarios.	Professor M cannot always quickly accommodate meeting requests they receive; students sometimes must wait over a week. They sometimes cross paths with students around campus and usually responds to emails in a timely manner, but students sometimes do not receive a reply for several days.
7-14	Inaccessible	Student rates faculty member they had in mind as much less	Professor B is rarely available to meet with students; they keep few office hours and do

	accessible than all scenarios.	not accommodate requests to meet at other times. The only time students see Professor B is during class. Professor B rarely responds to emails in a timely manner and sometimes does not respond at all.

Social Engagement Scale

Figure 4.38 presents the SES variable map with added interpretational information. The raw scale scores along the left side of the map range from 7 (indicating a selection of “much less engaged than X” for all items) to 35 (indicating a selection of “much more engaged than X” for all items). These raw scores allow for the division of the SES continuum into different score zones, for which the scenario items can serve as descriptions. These descriptions are illustrated in both Figure 4.38 and Table 4.37.

Students with the lowest perceptions of their faculty member’s social engagement scored from 7-14. These respondents were expected to find the faculty member that they rated *much less engaged* than what is depicted in all seven scenarios. Individuals in this range found their focal faculty member *much less engaged* than someone who: (a) seems annoyed during meetings with students and would clearly rather spend time on other activities, (b) does not acknowledge students around campus, and (c) sends emails that are disrespectful and do not address student questions or concerns. Very few respondents fall in this score zone.

Moving up the social engagement continuum, the few students scoring in the 15-21 range are expected to select *a little less engaged than X* for the items in their zone and *about the same as X* for items in the lowest zone. These students falling in this score range reported perceiving their focal faculty members as *a little less accessible* than a faculty member who (a) is not fully engaged in meetings with students and appears distracted (b) does not go out of their way to engage students in conversation around campus, and (c) writes emails that are unclear and may appear rude in tone.

Students scoring in the 22-28 range of the SES are predicted to select *about the same as X* for the items falling in their zone. These students perceived the faculty member they had in mind as about the same as an individual who (a) listens to student concerns during meetings but may not always fully engage, (b) acknowledges students inconsistently around campus, and (c) writes emails that are thorough and respectful. They are also predicted to select *a little more engaged than X* for items in lower zones.

Finally, respondents in the highest score zone (range of 29 to 35) were expected to select *a little more engaged than X* for the item in that zone and *much more engaged than X* for items in lower zones. This item, HHH, represents high perceptions of social engagement across all three facets. Substantively, this means that students in this score range perceived the faculty member they rated as a little more engaged than someone who is (a) clearly engaged during meetings, listens carefully, and asks questions, (b) seems happy to see students around campus, and (c) writes emails that always clearly address students' questions and concerns. As with the PAS, it is difficult to imagine exactly what someone who is more engaged than this looks like. This represents an area for further research, especially given the number of respondents who fell in this range.

Figure 4.38

Interpretational Social Engagement Scale Variable Map

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>  →                               Average Item Score = 5
   35             ##  +
                      |
                      |
   34                 |
                      +
                      |
                      |
   33             .# T|                  HHH: Professor K is clearly engaged during meetings
                      +                  with students; they listen to students carefully, ask
                      |                  questions, and engage in thoughtful discussion. They
                  ##  |                  seem happy to see students around campus and often stop
                      |                  to chat. Professor K’s responses to emails are always
                      +                  courteous and clearly address students’ questions or
   31      .######## S|T                 concerns.
                      |
            .#######  |  HHH
                      +
   29       .#######  |
   28     ##########  |  →                                    Average Item Score = 4
                     M|  HHM
   27          #####  +
               .####  |S                 MMH: Professor T listens to students’ concerns and
                      |                  questions during meetings, but does not always engage
                  .#  |                  in thoughtful conversation. They acknowledge students
   24          .####  +                  inconsistently when crossing paths around campus.
                 .## S|                  Professor T’s responses to students’ emails, however,
                      |                  are thorough and respectful.
                 .##  |  MMH
   21             .#  +M  →                                   Average Item Score = 3
                   .  |                  MML: Professor S is not always fully engaged during
                      |                  meetings with students; their attention sometimes
                   . T|  MMM             appears to be elsewhere. When they sees students around
   18              .  +                  campus, Professor S occasionally offers a wave of
                      |  LLM    MML      recognition, but does not go out of their way to engage
                      |                  in conversation. Professor S’s responses to emails are
   15                 |S                 often unclear and may appear rude in tone.
                      +
   14              .  |  LLL  →                               Average Item Score = 2
                      |
                      |
   13                 +                  LLL: Professor R often seems annoyed during meetings
                      |                  with students and would clearly rather spend time on
                      |                  other activities. They do not acknowledge students
                      |T                 around campus and their responses to emails are often
                   .  +                  disrespectful in tone, and do not clearly address
   11                 |                  student concerns or questions.
                      |
                      |
   9               .  +
   7           <less>|<freq>  →                               Average Item Score = 1
```

Table 4.37

Social Engagement Scale Score Interpretations

Score or Range	Perception Level	Description	Illustrative Scenario
35	Extremely engaged	Student rates faculty member they had in mind as much more engaged than all scenarios.	N/A
29-34	Very engaged	Student rates faculty member as a little more engaged than HHH and much more engaged than lower scenarios.	Professor K is clearly engaged during meetings with students; they listen to students carefully, ask questions, and engage in thoughtful discussion. They seem happy to see students around campus and often stop to chat. Professor K's responses to emails are always courteous and clearly address students' questions or concerns.
22-28	Moderately engaged	Student rates faculty member as about the same as HHM or MMH and as a little more engaged than lower scenarios.	Professor T listens to students' concerns and questions during meetings, but does not always engage in thoughtful conversation. They acknowledge students inconsistently when crossing paths around campus. Professor T's responses to students' emails, however, are thorough and respectful.
15-21	Moderately unengaged	Student rates faculty member as a little less engaged than MMM or MML and much less accessible than lower scenarios.	Professor S is not always fully engaged during meetings with students; their attention sometimes appears to be elsewhere. When they sees students around campus, Professor S

			occasionally offers a wave of recognition, but does not go out of their way to engage in conversation. Professor S's responses to emails are often unclear and may appear rude in tone.
7-14	Unengaged	Student rates faculty member they had in mind as much less engaged than all scenarios.	Professor R often seems annoyed during meetings with students and would clearly rather spend time on other activities. They do not acknowledge students around campus and their responses to emails are often disrespectful in tone, and do not clearly address student concerns or questions.

Secondary Analyses (Final Administration Data)

Descriptive Statistics

My dissertation's secondary research question concerns the degree to which students' perceptions of faculty availability outside of class as measured by the OCAS have relationships with students' actual participation in out-of-class communication. To capture these relationships, several additional items were administered to survey respondents following the scale items. These items asked students to provide information concerning the faculty member they had in mind while completing the OCAS with respect to the following:

- Estimate of the number of times they engaged in different forms of out-of-class communication,

- Whether or not they discussed specific topics outside of class, and
- Overall level of satisfaction with out-of-class communication.

Respondents were also asked to indicate their perceptions of the faculty member's gender and race/ethnicity, as well as department so that relationships concerning these variables could be explored in the future. However, they are not part of the analyses presented here.

Tables 4.38, 4.39, and 4.40 present frequencies for each of the topics noted above. Not all students completed these items, so sample sizes may differ from what was reported for the scale analyses in the previous section. Sample sizes for each item are presented within the tables.

Table 4.38

Types of Out-of-Class Communication Frequencies

Type of OCC	1 time (%)	2 times (%)	3 times (%)	4 times (%)	5 or more times (%)
Met during scheduled office hours (N = 184)	89 (48.4)	35 (19.0)	28 (15.2)	13 (7.1)	19 (10.3)
Met outside of scheduled office hours (N = 155)	110 (71.0)	17 (11.0)	16 (10.3)	3 (1.9)	9 (5.8)
Interacted informally around campus (N = 162)	87 (53.7)	35 (21.6)	18 (11.1)	10 (6.2)	12 (7.4)
Corresponded via email (N = 180)	35 (19.4)	30 (16.7)	41 (22.8)	26 (14.4)	48 (26.7)
Interacted immediately before or after class (N = 178)	42 (23.6)	42 (23.6)	36 (20.2)	19 (10.7)	39 (21.9)

Note: Percentages may not sum to 100 due to rounding.

Table 4.39

Out-of-Class Communication Topics Frequencies

OCC Topic	Discussed (%)	Not Discussed (%)
Grades (N = 187)	89 (47.6)	98 (52.4)
Questions about assignment directions or expectations (N = 187)	169 (90.4)	18 (9.6)
Academic advice (N = 187)	104 (55.6)	83 (44.4)
Career advice (N = 187)	43 (23.0)	144 (61.3)
Personal life (N = 187)	55 (29.4)	132 (70.6)
Letters of recommendation (N = 186)	12 (6.5)	174 (93.5)
Non-class related intellectual topics (N = 187)	50 (26.7)	137 (73.3)

Note: Percentages may not sum to 100 due to rounding.

Table 4.40

Out-of-Class Communication Satisfaction Frequencies

	N (%)
Very satisfied	113 (59.8)
Somewhat satisfied	61 (32.3)
Somewhat dissatisfied	11 (5.8)
Very dissatisfied	4 (2.1)

Note: Percentages may not sum to 100 due to rounding.

These tables provide a general picture of how sample students engaged in out-of-class communication with the faculty members they had in mind while completing the OCAS. Students within this sample appear to interact quite frequently with their focal faculty members, most frequently either via email or in-person immediately before/after class. By far the most common topic of these interactions is questions about assignment directions or expectations, although more than 50% of respondents also report discussing general academic advice with their focal faculty members. Students appear to be generally satisfied with this interaction, as more than 90% of respondents are either somewhat or very satisfied with out-of-class communication with their focal faculty members.

Scores on individual PAS and SES items were summed to create overall scores for each scale. The maximum possible score is 35 and the minimum possible score is 7. These are the scale scores used throughout the secondary analyses. Table 4.41 presents the means, standard deviations, maximum scores, and minimum scores for each scale. As expected from the Rasch analyses, the mean scores for both scales were quite high. Also as expected, the correlation between scores on the two scales is statistically significant ($r = .63$, $p < .001$). This relationship is also seen in a scatterplot of the two scale scores, shown in Figure 4.39. Substantively, the strong relationship between PAS and SES scores means that students' who had higher perceptions of a faculty member's physical accessibility also tended to have higher perceptions of the faculty member's social engagement. However, there are a few exceptions to this general trend. The two circled points in Figure 4.39 represent students who perceived the faculty members that they rated to be low in physical accessibility but high in social engagement.

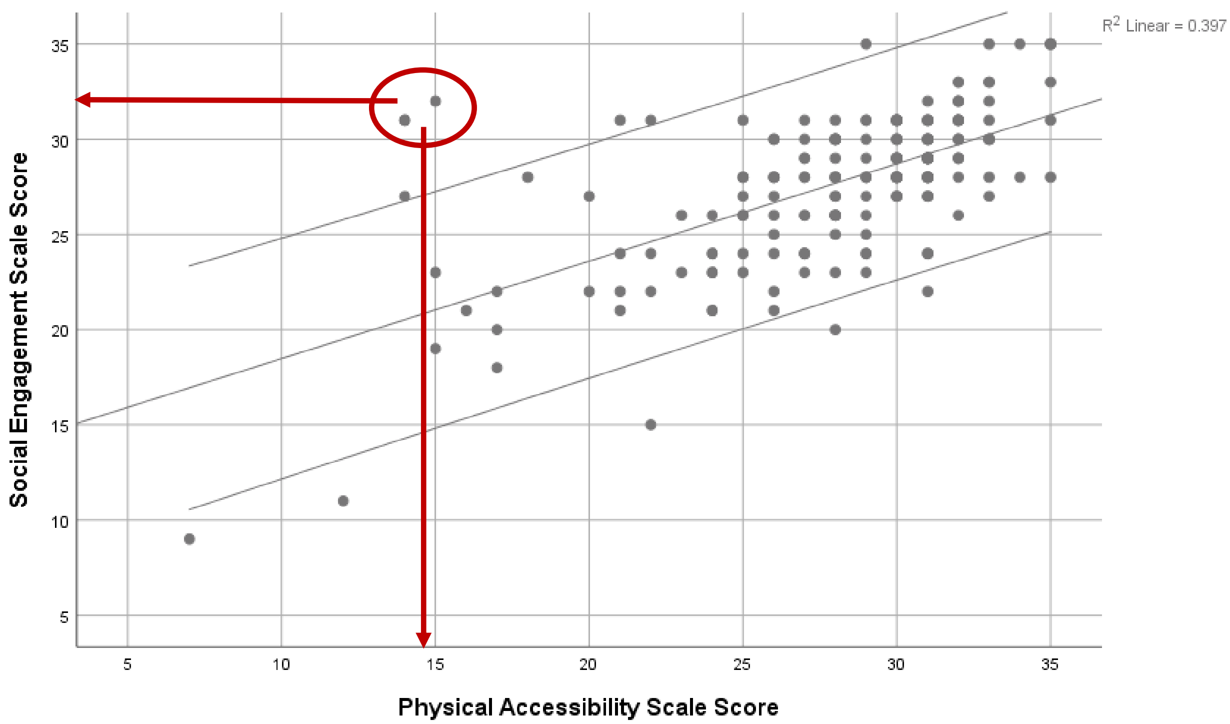
Table 4.41

Means and Standard Deviations for OCAS

	Mean	S.D.	Maximum	Minimum
Physical Accessibility Scale	27.88	4.89	35	7
Social Engagement Scale	27.65	3.97	35	9

Figure 4.39

Scatterplot of Physical Accessibility Scores and Social Engagement Scores



The information in Table 4.38 is complemented by histograms (Figures 4.40 and 4.41) illustrating the distributions of PAS and SES scale scores. Unsurprisingly, the distributions are negatively skewed by the relatively small number of low-scoring individuals.

Figure 4.40

Histogram of Physical Accessibility Scale Scores

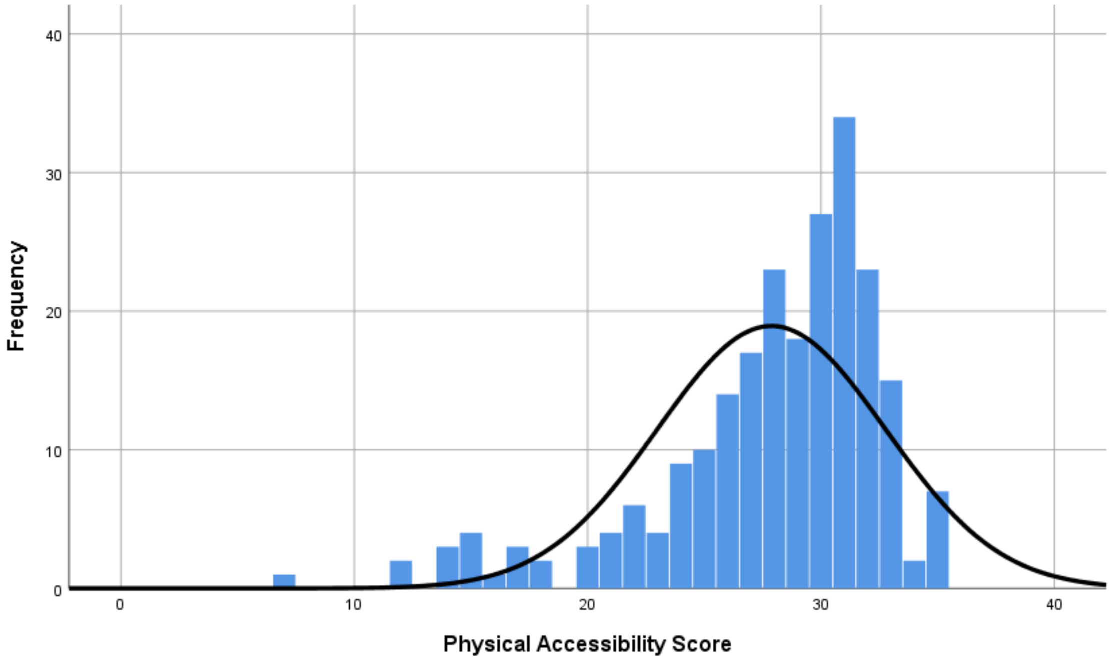
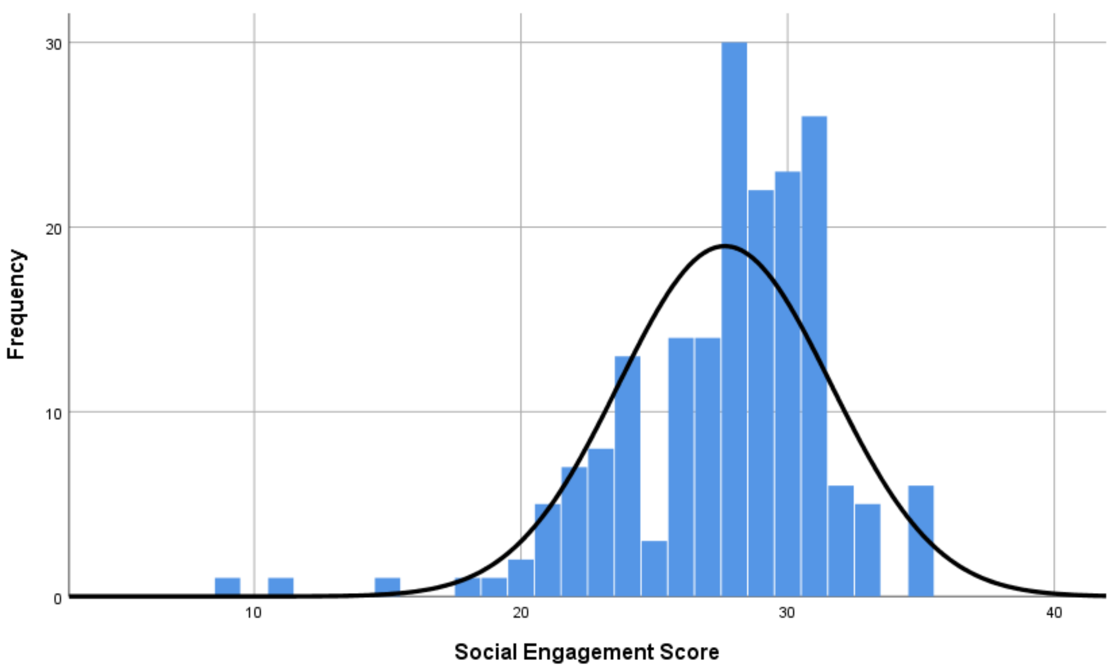


Figure 4.41

Histogram of Social Engagement Scale Scores



Finally, Table 4.42 presents the numbers of respondents falling into each of the score zones for each scale. These zones were demarcated on the final variable maps in Figures 37 and 38. This information is particularly important to consider when interpreting the relationships between OCAS scores and participation in OCC. For both scales, there are very few low-scoring individuals. This means that it will not be possible to definitively establish relationships between students' perceptions and various indicators of OCC participation. This is because there are so few low-scoring individuals in the sample that it is not possible to fully understand OCC participation of students who perceive their faculty members to how low physical accessibility or social engagement.

Table 4.42

Out-of-Class Availability Scales Score Zone Frequencies

Score Ranges	PAS Frequency (%) N = 232	SES Frequency (%) N = 189
7 – 14	6 (2.6)	2 (1.1)
15 – 21	17 (7.3)	10 (5.3)
22 – 28	83 (35.8)	89 (47.1)
29 – 35	126 (54.3)	88 (46.6)

Note: Only individuals with complete scale data are included in this table because individuals who did not complete all items could fall outside of these score ranges. This means that a small number of individuals who were included in the Rasch analyses are excluded here.

Relationships with OCAS Scores

Number of Out-of-Class Interactions

I examined the relationship between OCAS scores and the numbers of out-of-class interactions in which students engaged with their focal faculty members using scatterplots. OCAS scores appear on the *x*-axis and the number of interactions appear on the *y*-axis. These scatterplots appear in Appendix M.

The appearance of these scatterplots is affected by the nature of the data. Students reported meeting with their focal faculty members a discrete number of times. All responses of “5 or more” times are treated as “5” in these displays. This undoubtedly restricts the outcome variable ranges; however, it should still provide a general indication of the relationships between OCAS scores and the numbers of various out-of-class interactions. “5 or more” was used as a response option in the survey because it was deemed unlikely that students who met with faculty members that frequently would be able to provide an exact count of these meetings.

All of the relationships within these scatterplots are positive, but quite weak. As noted above, students in this sample scored quite high on the OCAS and also appeared to interact with their focal faculty members relatively often through various means. These concentrations of individuals at the high end of both sets of variables limits the extent to which relationships can be observed. As there are very few individuals with very low OCAS scores, from a substantive perspective, this means that sample students both perceive their focal faculty members to be available outside of class and also engage in out-of-class interaction with those faculty members. This is actually a positive finding, given the benefits of out-of-class communication that were presented in Chapter 2. However, it makes it difficult to determine exactly how perceptions of physical accessibility and social engagement are related to out-of-class interaction.

Out-of-Class Communication Topics

I investigated the relationship between OCAS scores and the discussion of various topics during out-of-class communication using logistic regression. I created two logistic regression models to predict whether or not students discussed each topic outside of class with the faculty member they rated: one using PAS score as a predictor and one using SES score as a predictor. The results of these models are presented in Appendix N.

In most cases, neither PAS nor SES scores were statistically significant predictors of a student's chance of discussing a particular topic during out-of-class communication with their focal faculty member. The only statistically significant findings concerned SES scores as a predictor for discussion of career advice and non-class related intellectual topics. Across the topics, SES scores tended to be stronger predictors than PAS scores, even when they did not reach statistical significance. This finding makes logical sense; it seems reasonable that a faculty member's behavior during out-of-class interaction might wield a strong influence on students' comfort raising particular topics.

As with the previous section concerning out-of-class interaction frequencies, it is possible that homogeneity of student responses explains the lack of significant findings. As has already been noted, many students were high-scoring on both the PAS and SES. This lack of variation makes it difficult to determine exactly how perceptions of physical accessibility and social engagement affect students' chances of discussing particular topics with faculty. This is simply because there are very few low-scoring students, so it is not possible to observe from these data what topics low-scoring students would indicate discussing.

Out-of-Class Communication Satisfaction

I used a Pearson correlational analysis to examine the relationship between OCAS scores and students' satisfaction with out-of-class communication with their focal faculty members. In contrast to the findings concerning frequency of different kinds of interactions, both OCAS scores had statistically significant correlations with students' out-of-class communication satisfaction (see Table 4.43). The correlation is slightly stronger for SES scores than for PAS scores. These relationships can also be examined visually in the scatterplots presented in Figures 4.42 and 4.43.

Table 4.43

Correlations of OCAS Scores with Out-of-Class Communication Satisfaction

	Pearson *r* with Satisfaction	Sig.
Physical Accessibility Scale Score	.410	<.001
Social Engagement Scale Score	.476	<.001

Figure 4.41

Physical Accessibility Scale Score and OCC Satisfaction

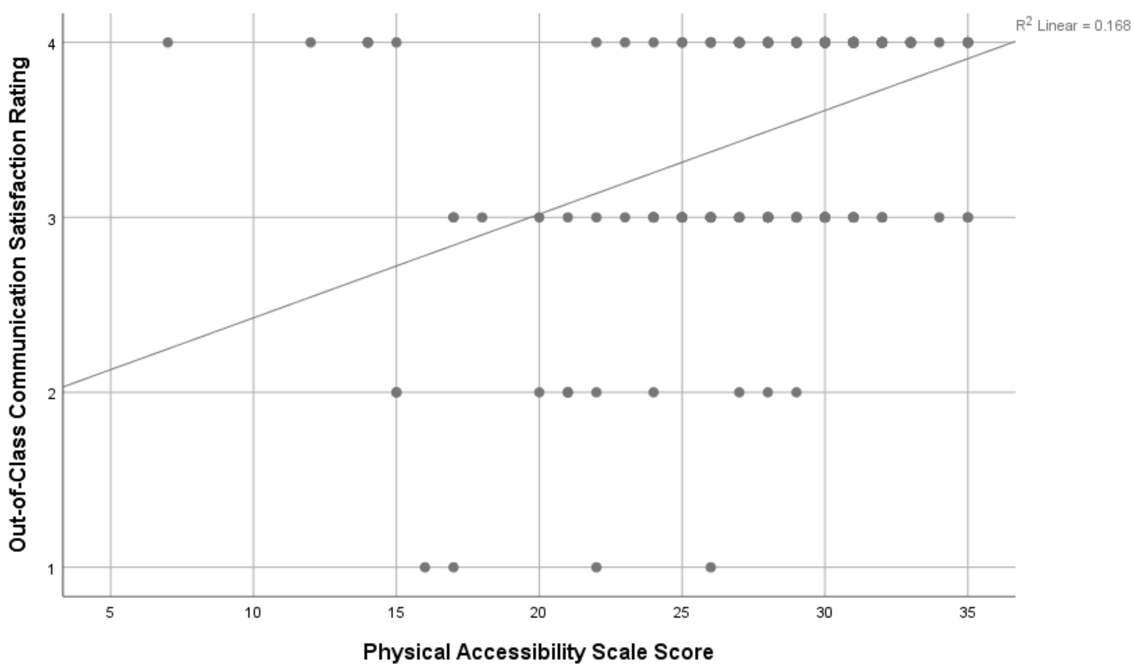
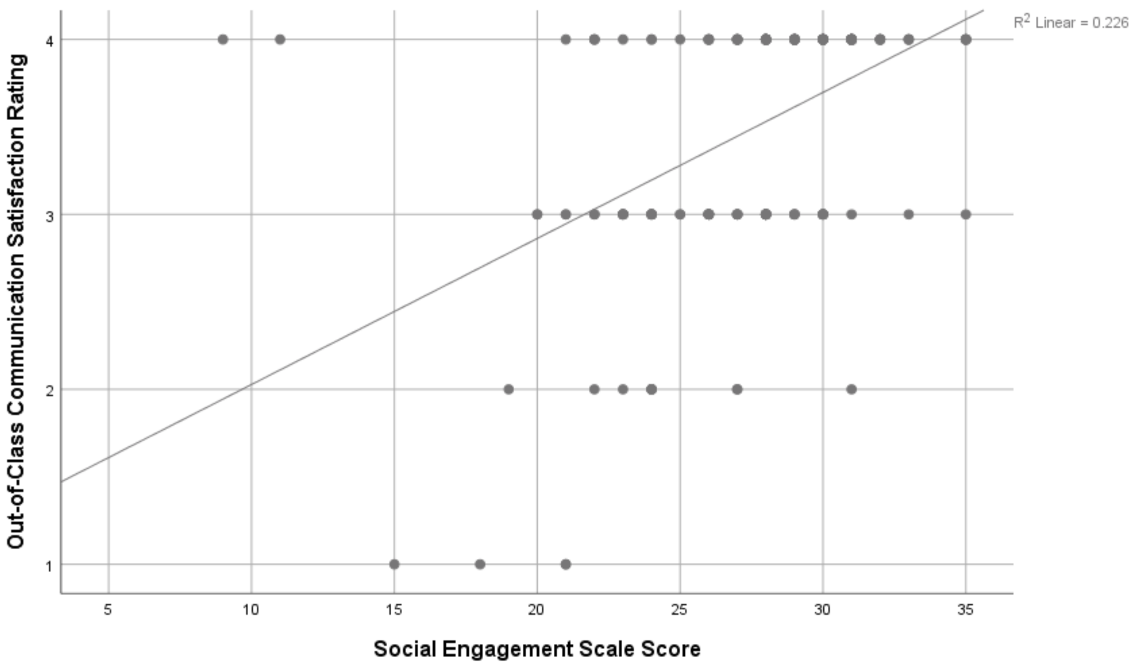


Figure 4.42

Social Engagement Scale Score and OCC Satisfaction



Chapter Summary

This chapter provided the results of my dissertation's primary and secondary research questions:

1. Can Rasch/Guttman Scenario scales provide valid and reliable measurement of student perceptions of faculty availability outside of class?
2. What is the relationship between scores from these scales and students' participation in out-of-class communication?

Three sets of analyses were presented to illustrate the evolution of the Out-of-Class Availability scales (OCAS): a pre-pilot, a pilot study, and a final administration. The measurement properties of the scales improved with each successive administration, as is demonstrated throughout this chapter. Future directions and intended uses for these scales are discussed in Chapter 5.

Results were also presented concerning the relationship between OCAS and students' participation in out-of-class communication with respect to frequency, the topics discussed, and satisfaction. Weak relationships were found for both frequency and topics; however, scores from both OCAS correlate strongly with students' satisfaction with out-of-class communication.

CHAPER FIVE: DISCUSSION

The previous four chapters introduced my dissertation's scope and purpose, summarized relevant literature, explicated my methodology, and presented results. The findings discussed in Chapter 4 illustrated how the Rasch/Guttman scenario (RGS) scale methodology (Ludlow et al., in press) was successfully applied to create an instrument measuring undergraduate students' perceptions of faculty availability outside of class. This chapter concludes my dissertation by outlining how this work contributes to existing research in both higher education and instrument development. I also discuss my work's limitations and directions for future research.

Overview of Results

My dissertation has addressed two research questions:

1. Can Rasch/Guttman Scenario scales provide valid and reliable measurement of student perceptions of faculty availability outside of class?
2. What is the relationship between scores from these scales and students' participation in out-of-class communication?

The RGS methodology was employed to create two scales, each capturing a dimension of students' perceptions of faculty availability outside of class. These scales are the Physical Accessibility Scale (PAS) and the Social Engagement Scale (SES), and collectively they form the Out-of-Class Availability Scales (OCAS).

After a thorough review of existing instruments and lengthy item development process, I began my empirical analysis with a small pre-pilot study to confirm that the physical accessibility and social engagement constructs operated as I had hypothesized with respect to the three facets of arranged meetings, chance encounters, and email.

These results were promising; items deliberately constructed to represent high levels of accessibility or engagement were more difficult for respondents to endorse (i.e., to select that their focal faculty member was more accessible/engaged than the scenario description) and items constructed to reflect low levels of accessibility or engagement were less difficult for respondents to endorse. However, results were mixed in the middle areas of the continuum, and one SES item was much easier than intended. These results were used to inform item revisions, which were then evaluated in the pilot study. The pilot study results provided further empirical support for the hypothesized construct continua, and indicated success of item revisions. Several additional revisions were made prior to the final scale administration. This process ultimately resulted in two seven-item scales that can be used to evaluate students' perceptions of faculty availability outside of class. The PAS and SES scores were highly correlated with each other, which is expected, given that they are both conceptualized as dimensions of students' perceptions of faculty availability outside of class.

Findings were somewhat less clear with respect to my secondary research question. Both PAS and SES scores displayed positive, but weak relationships with the number of times students reported interacting outside of class with the faculty members that they rated. PAS and SES scores were also largely insignificant predictors of students' chances of discussing particular topics with faculty members outside of class. However, both scale scores had a statistically significant, positive correlation with students' satisfaction with out-of-class communication with the faculty member.

Most students in the final sample had high scores on both the PAS and SES, which may account for the observed relationships between scale scores and OCC participation. Because so few students scored at the low end of the scale, it is not possible to discern a clear picture of how students who do not perceive a faculty member to be physically accessible or socially engaged participate in out-of-class communication. Thus, the relationship between students' perceptions as measured by the OCAS and their actual participation could not be firmly established. Were more low-scoring students present in the sample, it is possible that more statistically significant relationships would have been observed, and that the correlation between satisfaction and OCAS scores would be even stronger.

Discussion of Results

In Chapter 2 I discussed the limitations of existing instruments related to OCC between undergraduate students and faculty. These fell into two broad categories: limitations related to construct definition and limitations related to measurement approaches. From a content perspective, existing instruments tended to focus solely on frequency—in other words, they asked students to indicate *how often* they engaged in OCC. While useful for establishing the importance of OCC in influencing undergraduate outcomes, this conceptualization does little to reveal *why* students engage in OCC. Some instruments also defined OCC quite narrowly, perhaps focusing solely on interaction that occurs face-to-face and failing to encompass other spaces in which faculty and students interact. From a measurement perspective, these instruments were developed almost exclusively with Likert-type items in the classical test theory (CTT). This form of scale development treats items as replications and yields results that are completely sample-dependent, meaning that scores from CTT scales are lacking in clear interpretations.

The OCAS address these limitations through the fusion of Guttman facet theory design and the Rasch measurement paradigm. Application of facet theory design principles led to the identification of different *facets*, or components, of students' perceptions of faculty availability outside of class. These facets describe the different spaces and places in which students and faculty may interact, suggesting that they all come together to shape the unidimensional construct of students' perceptions. These facets are situated within two dimensions: physical accessibility and social engagement. Altogether, these facets and dimensions come together in the form of two scales, each of which has clearly interpretable scores because of the Rasch measurement framework. These interpretations were illustrated in Chapter 4 through Figures 4.37 and 4.38 (pp. 209 & 213).

For the PAS, the small number of students scoring in the 7-14 range (N = 6) are likely to find the faculty member they rated as *much less accessible than* what is depicted in all seven scenarios. Students scoring 14-21 (N = 17) are likely to find the faculty member they considered as *a little less accessible* than someone who 1) cannot always accommodate meeting requests and sometimes keeps students waiting for over a week; 2) sometimes crosses paths with students around campus; and 3) usually responds to emails in a timely manner, but sometimes does not respond for several days. Proceeding through the continuum, students with scores in the 22-28 range (N = 83) are likely to rate their focal faculty member as *about the same as* someone who 1) holds regular office hours, but will not meet students outside of them; 2) is only seen outside of class occasionally; and 3) responds to emails within one weekday. They are also predicted to find their faculty member *a little more accessible* than what is described in items lower on the scale. Finally, students with scores in the 29-35 range (N = 126) were likely to find the faculty member that they rated *a little more accessible* than someone who 1) holds regular office hours and meets outside of them when necessary; 2) seems to always be on campus and is often seen by students; and 3) always responds to emails in a timely manner. These students were also likely to find their focal faculty member *much more accessible* than all other scenario representations.

The SES scores can be interpreted in a similar fashion. At the low end of the continuum, students who scored 7-14 (N = 2) are expected to rate the faculty member they had in mind as *much less engaged than* what is depicted in all seven scenarios. Students scoring from 7-14 (N = 10) likely found their faculty member a little less engaged than someone who 1) is not fully engaged in meetings with students and appears distracted; 2) does not go out of their way to engage students in conversation around campus; and 3) writes emails that are unclear and may appear rude in tone. Students who scored in the 22-28 range (N = 89) are predicted to perceive their focal faculty member as *about the same* as someone who 1) listens to student concerns during meetings but may not always fully engage; 2) acknowledges students inconsistently around campus; and 3) writes emails that are thorough and respectful. In addition, they are predicted to find their faculty member *a little more engaged* than what is depicted in scenarios further down in the construct continuum. At the top of the continuum, students scoring from 29-35 (N = 88) were predicted to find the faculty member they rated as *a little more engaged* than someone who is 1) clearly engaged during meetings, listens carefully, and asks questions; 2) seems happy to see students around campus; and 3) writes emails that always clearly address students' questions and concerns. They also likely found the faculty member they rated as much more engaged than what is depicted in all other scenarios.

These scores provide information that is more useful than a count of frequencies or a Likert scale score. Consider a faculty member who administers the OCAS to her students. After administration, she would receive more than a decontextualized mean rating. The mean scale scores would fall within a score range that has a qualitative interpretation in the form of the scenario items. Not only would this be useful for describing current status, but it could also reveal ways in which the faculty member might work to alter students' perceptions. This might also reveal divergence between students' perceptions of the faculty member and her actual behaviors or intentions (more on this below). In addition to a mean score, the faculty member could inspect her variable map to see the distribution of her students' along the physical accessibility and social engagement continua. Because the Rasch estimation procedure will still work with relatively small sample sizes, use of the OCAS is an option for faculty teaching many different kinds of classes.

At a higher level, department heads or institutional research offices might also make use of the OCAS to get a broad picture of students' perceptions regarding faculty availability outside of class using a sampling method similar to what I employed in my dissertation. Ideally, a representative sample of students could be recruited and asked to complete the scales thinking about the individual with whom they most recently had class. Again, rather than a decontextualized mean score, the interested party would receive qualitative interpretations, along with a variable map illustrating the comparability of respondents and items.

In addition to the score interpretability that comes from the RGS development process, the OCAS are also useful to higher education researchers because of the OCC-related construct they are designed to capture. The OCAS measure students' *perceptions* of faculty availability outside of class. This stands in contrast to other instruments (e.g., Knapp & Martin, 2002; Knapp, 2005), which focused on how *often* students engaged in OCC. Students' perceptions of availability are more telling for the purposes of faculty or department evaluation than frequency. Even if a student does not actually engage in OCC with a particular faculty member, ideally they would still perceive that faculty member as available for doing so.

Focusing on students' perceptions of availability can also help disentangle how engagement in OCC differs across various student groups. As highlighted in Chapter 2, some research indicates (e.g., Cole, 2010; Kim & Sax, 2011) that students may engage in or benefit from OCC differently across racial and gender groups. Examining variation in perceptions of faculty availability across these groups may provide additional context for these findings.

Finally, understanding students' perceptions of faculty availability may shed light on why OCC does not always occur frequently for students, despite its positive benefits (Cotten & Wilson, 2006; Denzine & Pulos, 2000). The weak relationships between OCAS scores OCC participation in my dissertation may have been attributable to sampling particularities; whether or not students' perceptions of availability influences OCC should be further explored.

Limitations

Despite these promising findings and contributions, my dissertation has several limitations. One of these is the reliance on a single institution for the collection of pilot and final administration data. Not only do all respondents come from a single institution, but they also come from a selective Jesuit institution that purports to emphasize education of the whole person. It is certainly possible (and plausible) that these students are not representative of the general undergraduate population in the United States. Students in my sample (and at this institution in general) may well find their faculty members to be much more accessible outside of class than would students at larger, less selective institutions that do not actively emphasize the education of the whole person. This may be because of the institution itself or because of specific characteristics of students attending the institution; it is not possible to know for sure. The same general premise applies to faculty: it may be that faculty working at this particular institution place a high value on being available to students outside of class.

The lack of representativeness of this sample may have impacted my dissertation results in a substantial way. As was noted repeatedly throughout Chapter 4 for both the PAS and the SES, the distribution of respondents was consistently concentrated at the higher end of the construct continua, indicating that students found the faculty members that they had in mind to be most similar to the descriptions in the more difficult scenarios. This is a positive finding from the perspective of the institution because it indicates that students do find their instructors to be both physically accessible and socially engaged outside of class. However, from a measurement perspective, it means that the scale items and individuals are not particularly well-aligned. These generally high scores also impacted my ability to ascertain the relationship between students' perceptions of accessibility, engagement, and participation in OCC. Because so few students had low scores, I do not have a clear idea of how students engage in OCC with a faculty member whom they do not find to be available outside of class.

Characteristics of the sample also may have affected the results of my secondary research question concerning the relationship between OCAS scores and students' self-reported participation in OCC. Students generally reported interacting a great deal with their focal faculty members outside of class. This finding, coupled with the high OCAS scores, may have obscured relationships between perceptions of availability and participation in OCC, which may account for the largely insignificant relationships that OCAS scores had with various forms of OCC.

Another limitation with respect to my samples are size and response rate. This was of less concern for the pre-pilot because of its truly exploratory nature; however, the response rates for both my pilot study and final administration were quite low. 1,500 students were invited to participate in the pilot study; after accounting for missing data, I had 64 responses for the PAS (response rate of 4.2%) and 53 responses for the SES (response rate of 3.5%). 3,000 students were invited to participate in the final administration; after accounting for missing data, I had 237 responses for the PAS (response rate of 7.9%) and 192 responses for the SES (response rate of 6.4%). While the sample sizes in both cases were sufficient to complete (and, for the final administration, have some confidence in) the Rasch analyses, they are not likely to be fully representative of the larger population from which the sample was drawn.

One final limitation that merits consideration is that the OCAS measure *students' perceptions* of faculty availability. It has been noted several times that these perceptions may differ from how faculty perceive themselves, or even from what one might observe if they were to track a faculty members' accessibility for or engagement in OCC. Some might argue that this is problematic, and that the prioritizing of students' perceptions serves to further push a consumer-oriented mindset into higher education. However, divergence between students' perceptions of a faculty member and that faculty member's perceptions of him/herself may actually serve as a useful point of exploration. This possibility is discussed in more detail in the following section.

Future Directions

My dissertation has provided promising evidence for the use of Rasch/Guttman scenario scales to measure undergraduate students' perceptions of faculty availability outside of class. There are several ways in which to build on this initial work. First, some scenario items could be further revised to reduce redundancy and increase spread in the middle areas of the scales. For example, within the SES, Items MMM, MML, and LLM were clustered quite close together (MML and LLM were nearly equivalent in difficulty). There is also a relatively large gap in the continuum between Items MMH and HHM. These items could be revised to ensure that the SES conforms more closely to the principles of Rasch measurement and reflects a more ladder-like progression of the social engagement construct.

It would also be helpful to further explore the high ends of the physical accessibility and social engagement continua. Although the high-scoring nature of my pilot and final respondents may be in part due to the sample, the pre-pilot respondents (who came from another institution) also scored quite highly on the first version of the OCAS. As noted in Chapter 4, it is somewhat difficult to imagine a faculty member who is perceived as more accessible or engaged than what is described in the scales' most difficult scenarios. Because my dissertation focus groups were focused on verifying the general structure of the scales' continua, facet identification, and diagnosing problematic scenario wording, this is not an issue that was explored in detail. A subsequent round of focus groups focused explicitly on what an extremely accessible or engaged faculty member looks like could assist in either revising the existing high-level items or the addition of new items reflecting even higher perceptions.

After revision, it would be also beneficial to administer the OCAS to a student sample at a different institution. I noted in the limitations how the exclusive use of students from a particular institution may have impacted my results; obtaining a sample from a different institution could verify these speculations. Beyond this, administration of scales to multiple samples is an important validation activity. OCAS results conforming to Rasch measurement principles obtained from multiple samples would increase overall confidence in the validity and reliability of the scales.

Additional validity support for the OCAS might also be obtained through application of the Many-Facet Rasch Model (MFRM) (Linacre, 1989). Unlike the rating scale model applied here, which treats all facets together as components of a unitary construct, the MFRM tests the influence of individual facets upon scale scores, and adjusts those scores accordingly (Eckes, 2015). Because of this, the MFRM could be used to test the hypothesis that all three OCAS facets (arranged meetings, chance encounters, and email) actually do influence students' perceptions in a comparable way. This seems like a worthwhile undertaking because of the particular OCAS items that were disordered or displayed misfit in the final calibration. For example, Item HHM in the PAS was not in its intended difficulty order and had an INFIT statistic of 1.3. This scenario described a faculty member who was high on the arranged meetings and chance encounters facet and medium on the email facet. This item had a lower difficulty than MMH, which depicted a medium level on the arranged meetings and chance encounters facets but a high level on the email facet. This reversal may indicate that the email facet may be more influential in students' responses than the "in-person" facets. This seems not only possible, but also plausible given the increased use of email for OCC in recent years. This is something that could be tested empirically using the MFRM.

In addition to OCAS revisions, there are other areas of future research that might be explored. As noted in the Limitations section, the OCAS measure students' perceptions of faculty availability outside of class, which may differ from faculty members' own perceptions or behaviors. Once the student version has been sufficiently validated, a faculty version of the OCAS might be developed, where faculty could rate themselves in relation to the scenarios. Faculty scores could be considered in conjunction with Rasch results obtained from students to assess congruence or divergence. Because OCAS scale scores have qualitative interpretations, any observed differences between student and faculty descriptions could actually be described in words rather than simply quantified.

In addition to further OCAS-specific research, my dissertation also raises importance questions about the utility of RGS scales for faculty evaluation. I hypothesize that the interpretability of RGS scale scores makes them much more useful for informing faculty development and evaluation than typical Likert-type items. For example, a department chair might use the qualitative descriptions accompanying a faculty member's score to jumpstart conversations about professional growth. Additionally, students may find such items easier to respond to than traditional course evaluation items (despite the greater reading load) because of the specific behaviors depicted in the scenarios. At present, these are speculations; I hope to put the OCAS to use in a faculty evaluation context so that they can be verified with empirical data.

Implications and Conclusion

My dissertation has illustrated how Rasch/Guttman scenario scales can be used to measure undergraduate students' perceptions of faculty availability outside of class. The procedures and results presented have implications for research in both higher education and instrument development. From a content perspective, the OCAS provide a new way of framing student-faculty interaction outside of class. They also suggest that RGS scales may have other applications in higher education. Methodologically, the OCAS are one of the few currently existing RGS scales, and the documentation of their development process is instructive for future scale developers.

Interacting with faculty outside of class is associated with many positive outcomes for undergraduate students, including higher grades in specific courses (Guerro & Rod, 2013; Micari & Pazos, 2012), intrinsic motivation (Komarraju et al., 2010), and aspirations for further education (Hanson et al., 2016; Trolian & Parker, 2017). In addition to these broad trends, OCC may be particularly impactful for specific student groups, such as international or gender-variant students (BrckaLorenz et al., 2017; Glass et al., 2015). Because of these positive outcomes, higher education institutions have a clear interest in promoting OCC. The OCAS can aid in this venture by providing an indicator of students' perceptions that faculty are *available* for interaction outside of class. The OCAS attend to these perceptions in a multifaceted fashion, attending to both physical accessibility and social engagement across a variety of spaces and places. Scores from the OCAS can be meaningfully interpreted in order to inform changes in faculty practice or department/institutional strategy as appropriate.

Methodologically, my dissertation provides a clear and detailed account of the development of an RGS scale application. As noted above, few of these scales currently exist; however, interest in the methodology is growing (e.g., Antipikina & Ludlow, in press; Chang et al., 2019; Ludlow et al., in press). The explicitly articulated argument for the use of Guttman facet theory design and Rasch measurement procedures (see Chapter 2) strengthens the methodological foundation of these scales. This contribution, coupled with extensive documentation of the item development procedures (see Chapter 3), helps solidify the RGS methodology with the scale development literature. My work presented in this dissertation advances the RGS methodology in several ways. First, it is the only application where respondents are ultimately tasked with rating another person (i.e., students rating faculty) against a scenario item rather than themselves. The complications this introduces were discussed in Chapter 3. Second, none of the published RGS applications make the explicit theoretical argument for the integration of Rasch principles and facet design that I advance in the previous section. This contribution strengthens the theoretical foundation of these existing scales and those that will be developed in the future.

The successful creation of the OCAS not only provides a new lens for thinking about student-faculty interaction outside of class, but also opens the door for the use of RGS scales to measure additional constructs of interest to postsecondary institutions. The utility of scale scores with clear interpretations is not confined to the realm of out-of-class communication. The OCAS are but a first step in the use of this promising scale development methodology in the realm of higher education.

References

Anaya, G. & Cole, D. (2001). Latina/o student achievement: Exploring the influence of student-faculty interaction on college grades. *Journal of College Student Development, 42*(1), 3-14.

Andrich, D. (2004). Controversy and the Rasch model: A characteristics of incompatible paradigms? *Medical Care, 42*(1), 7-16.

Andrich. D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 561-573.

Antipikina, I. & Ludlow, L. (in press). Measuring parental involvement as a holistic concept using Rasch/Guttman scenario scales. *Journal of Psychoeducational Assessment.*

Astin, A. (1993). What matters in college? *Liberal Education, 79*(4), 4-15.

Atamian, R. & DeMoville, W. (1998). Office hours-- none an e-mail experiment. *College Teaching, 46*(1), 31-35.

Aylor, B. & Oppliger, P. (2003). Out-of-class communication and student perceptions of instructor humor orientation and socio-communicative style. *Communication Education, 52*(2). 122-134.

Bippus, A. M., Brooks, C. F., Plax, T. G., & Kearney, P. (2001). Students' perceptions of part-time and tenured/tenure-track faculty: Accessibility, mentoring, and extra-class communication. *Journal of the Association for Communication Administration, 30*, 13–23.

Bippus, A., Kearney, P., Plax, T., & Brooks, C. (2003). Teacher access and mentoring abilities: predicting the outcome value of extra class communication. *Journal of Applied Communication Research 31*(3) 260-275.

Bjorklund, S., Parente, J. & Sathianathan. (2004). Effects of faculty interaction and feedback on student skills. *Journal of Engineering Education*, *93*(2) 153-160.

Bock, R. (1997). A brief history of item response theory. *Educational Measurement: Issues and Practice, 16*(4), 21-33.

Borg, I., & Shye, S. (1995). *Facet theory*. Sage.

BrckaLorenz, A., Garvey, J., Hurtado, S., & Latopolski, K. (2017). High-impact practices and student-faculty interactions for gender-variant students. *Journal of Diversity in Higher Education, 10*(4), 350-365.

Canter, D. (1985). How to be a facet researcher. In D. Canter (Ed.), *Facet theory approaches to social research* (pp. 265-275). Springer-Verlag.

Chambliss, D. & Takacs, C. (2014). *How college works.* Harvard University Press.

Chang, W-C., Ludlow, L. H., Grudnoff, L., Ell, F., Haigh, M, Hill, M., & Cochran-Smith, M.

(2019). Measuring the complexity of teaching practice for equity: Development of a scenario-format scale. *Teaching and Teacher Education, 82,* 69-85.

Chen, W., Lenderking, W., Jin, Y., Wyrwich, K., Gelhorn, H., & Revicki, D. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research, 23,* 485-493.

Clark, R., Walker, M., & Keith, S. (2002). Experimentally assessing the student impacts of out-of-class communication: Office visits and the student experience. *Journal of College Student Development, 43*(3), 824-837.

Cohen, E. (2018). Gendered styles of student-faculty interaction among college students. *Social Science Research, 75*, 117-129.

Cokley, K., Komarraju, M., Patel, N., Castillon, J., Rosales, R., Pickett, R., Piedrahita, S., Ravitch, J., Pang, L. (2004). Construction and initial validation of the student-professor interaction scale (SPIS). *The College Student Affairs Journal, 24*(1), 33-51.

Cokley, K., Rosales, R., Kommaraju, M., Shen, F., Pickett, R., & Patel, N., (2007). Assessment of quality of student-faculty interactions. *Journal of the Professoriate, 1*(2), 53-67.

Cole, D. (2010). The effects of student-faculty interactions on minority students' college grades: Differences between aggregated and disaggregated data. *Journal of the Professoriate, 3*(2), 137-160.

Cotten, S. & Wilson, B. (2006). Student-faculty interactions: Dynamics and determinants. *Higher Education, 51*, 487-519.

Cox, B. (2011). A developmental typology of faculty-student interaction outside the classroom. *New Directions for Institutional Research*, *S1*, 49-66.

Cox, B., McIntosh, K., Terenzini, P., Reason, R., Lutovsky Quaye, B. (2010). Pedagogical signs of faculty approachability: factors shaping faculty-student interaction outside the classroom. *Research in Higher Education*, 51(8), 767-788.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory.* Holt, Rinehart, and Winston, Inc.

de Ayala, R. J. (2009). *The theory and practice of item response theory.* The Guilford Press.

de Souza, B., de Souza, F., Roazzi, A., & Lula, A. (2014). The lack of a treatment of uncertainty in facet theory: A relevant problem? In *Proceedings of the 14th Facet Theory Conference* (pp. 60-72). Recife: Brazil.

Denzine, G. & Pulos, S. (2000). College students' perceptions of faculty approachability. *Education Research Quarterly, 24*(1), 56-66.

Dobransky, N. & Frymier, A. (2004). Developing teacher-student relationships through out of class communication. *Communication Quarterly, 52*(3), 211-223.

Endo, J., & Harpel, R. (1982). The effect of student-faculty interaction on students' educational outcomes. *Research in Higher Education, 16*(2). 115-138.

Engelhard, G. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. Routledge.

Fabrigar, L., & Wegener, D. (2012). *Exploratory factor analysis.* Oxford University Press.

Faranda, W. (2015). The effects of instructor service performance, immediacy, and trust on student-faculty out-of-class communication. *Marketing Education Review, 25*(2), 83-97.

Feldman, Kenneth A. (1997). Identifying exemplary teaching and teaching: Evidence from student ratings. In Perry, Raymond & Smart, John C., *The scholarship of* *teaching and learning in higher education: An evidence-based perspective* (pp. 93-129). Springer.

Foral, P. A., Turner, P. D., Monaghan, M. S., Walters, R. W., Merkel, J. J., Lipschultz, J. H., & Lenz, T. L. (2010). Faculty and student expectations and perceptions of e-mail communication in a campus and distance doctor of pharmacy program. *American Journal of Pharmaceutical Education, 74*(10), 191.

Fusani, D. (1994). “Extra-class” communication: frequency, immediacy, self-disclosure, and satisfaction in student-faculty interaction outside the classroom. *Journal of Applied Communication Research, 22,* 232-255.

Glass, C., & Kociolek, E., Wongtrirat, R., Lynch, R. J., Cong, S. (2015). Uneven experiences: The impact of student-faculty interactions on international students’ sense of belonging. *Journal of International Students, 5*(4), 353-367.

Goldman, Z., Goodboy, A., & Bolkan, S. (2016). A meta-analytical review of students’ out-of-class communication and learning effects. *Communication Quarterly, 54*(4), 476-493.

Goodboy, A., Booth-Butterfield, M., Bolkan, S., & Griffin. (2015). The role of instructor humor and students’ educational orientations in student learning extra effort, participation, and out-of-class communication. *Communication Quarterly, 61*(1), 44-61.

Guerrero, M. & Rod, A. (2013). Engaging in office hours: A study of student-faculty interaction and academic performance. *Journal of Political Science Education, 9*, 403-416.

Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review, 9*(2), 139-150.

Guttman, L. (1954). An outline of some new methodology for social research. *Public Opinion Quarterly, 18,* 395-404.

Guttman, L. (1959) Introduction to facet design and analysis. In *Proceedings of the Fifteenth International Congress of Psychology* (pp. 130-132). Amsterdam: North Holland.

Guttman, R., & Greenbaum, C. W. (1998). Facet theory: Its development and current status. *European Psychologist, 3*(1), 13-36.

Hackett, P. (2014). *Facet theory and the mapping sentence*. Palgrave Macmillan.

Hambleton, R. & Cook, L. (1977). *Latent trait models and their use in analysis of educational test data. Journal of Educational Measurement, 14*(2), 75-96.

Hanson, J., Paulsen, M., & Pascarella, E. (2016). Understanding graduate school aspirations: the effect of good teaching practices. *Higher Education, 71*, 735-752.

Jaasma, Marjorie A., & Koper, Randall J. (1999). The relationship of student-faculty out of class communication to instructor immediacy and trust and to student motivation. *Communication Education, 48,* 41-47.

Jack, A. (2016). (No) harm in asking: Class, acquired cultural capital, and academic engagement at an elite university. *Sociology of Education, 89*(1), 1-19.

Kelly, L., Duran, R., & Zoltan, J. (2001). The effect of reticence on college students' use of electronic mail to communicate with faculty. *Communication Education, 50*(2), 170-176.

Kelly, L., Keaten, J., & Finch, C. (2004). Reticent and non-reticent college students' preferred communication channels for interacting with faculty. *Communication Research Reports, 21*(2), 197-209.

Khan, S., Shah, A., & Ahmad, S. (2015). The role of out-of-class communication in instructor's verbal/non-verbal behavior, trust, and student motivation. *Business & Economic Review, 7*(1), 81-100.

Kim, Y. K. & Sax, L. (2014). The effects of student-faculty interaction on academic self-concept: Does academic major matter? *Research on Higher Education, 55*, 780-809.

Kim, Y. K. & Sax, L. (2011). Are the effects of student-faculty interaction dependent on academic major? An examination using multilevel modeling. *Research on Higher Education, 52*, 589-615.

Kim, Y. K. & Sax, L. (2007). Different patterns of student-faculty interaction in research universities: An analysis by student gender, race, SES, and first-generation status. Center for Studies in Higher Education, University of California, Berkeley.

Knapp, J. (2005). *Exploring out-of-class communication: The development and testing of a measure* (master's thesis). West Virginia University, Morgantown, WV.

Knapp, J. & Martin, M. (2002). Out-of-class communication: The development and testing of a measure. Paper presented at the annual meeting of the Eastern Communication Association, New York, NY.

Komarraju, M., Musulkin, S., & Bhattacharya. (2010). Role of student-faculty interactions in developing college students' academic self-concept, motivation, and achievement. *Journal of College Student Development, 51*(3), 332-342.

Kuh, G. (2008). *High-impact educational practices: What they are, who has access to them, and why they matter.* Association of American Colleges and Universities.

Kuh, G. (1995). The other curriculum: Out-of-class experiences associated with student learning and personal development. *The Journal of Higher Education, 66*(2), 123-155.

Kuh, G. & Hu, S. (2001). The effects of student-faculty interaction in the 1990s. *The Review of Higher Education, 24*(3), 309-332.

Kuh, G., Hu, S., & Vesper, N. (2000). “They shall be known by what they do”: An activities-based typology of college students. *Journal of College Student Development, 41,* 228-244.

Kuh, G., Kinzie, J., Schuh, J., & Whitt, E. (2005). *Student success in college: Creating conditions that matter.* Jossey-Bass.

Laird, T. & Cruce, T. (2009). Individual and environmental effects of part-time enrollment status on student-faculty interaction and self-reported gains. *The Journal of Higher Education, 80*(3), 290-314.

Lechuga, V. M. (2011). Faculty-graduate student mentoring relationships: Mentors’ perceived roles and responsibilities. *Higher Education, 62,* 757-771.

Li, L. & Pitts, J. (2009). Does it really matter? Using virtual office hours to enhance student-faculty interaction. *Journal of Information Systems Education, 20*(2), 175-185.

Linacre, J. M. (2018). Winsteps® Rasch measurement computer program. Beaverton, Oregon: Winsteps.com

Linacre, J. (2018). Winsteps help for Rasch analysis. Retrieved from: https://www.winsteps.com/winman/

Linacre, J. (2002). Optimizing rating scale effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Linacre, J. (2002). What do infit, outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 878.

Linacre, J. (1998). Rasch analysis first or factor analysis first? *Rasch Measurement Transactions, 11*(4), 603.

Linacre, J. (1994). Sample size and item calibration [or person measure] stability. *Rasch Measurement Transactions, 7*(4), 328.

Loes, C. N, Saichaie, K., Padgett, R., & Pascarella, E. (2012). The effect of teacher behaviors on students' inclination to inquire and lifelong learning. *International Journal for the Scholarship of Teaching and Learning, 6*(2), 1-20.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. LEA Publishers.

Ludlow, L., Reynolds K. A., Baez-Cruz, M., & Chang, W. C. (in press). Enhancing the interpretation of scores through Rasch-based scenario-style items. In A. G. Harbaugh (Ed.), *Basic elements of survey research in education: Problems your advisor never told you about.*

Ludlow, L., Anghel, E., Szendey, O., O'Keefe, T., Howell, B., Matz-Costa, C., & Braun, H. (in press). The Boston College living a life of meaning and purpose (BC-LAMP) portfolio: An application of Rasch/Guttman scenario methodology. *Journal of Applied Measurement.*

Ludlow, L., Matz-Costa, C., & Klein, K. (2019). Enhancement and validation of the productive engagement portfolio-scenario (PEP-S8) scales. *Measurement and Evaluation in Counseling and Development, 52*(1), 15-37.

Ludlow, L., Matz-Costa, C., Johnson, C., Brown, M., Besen, E., & James, J. (2014). Measuring engagement in later life activities: Rasch-based scenario scales for work, caregiving, informal helping, and volunteering. *Measurement in Evaluation in Counseling and Development, 47*(2), 127-149.

Ludlow, L. & Haley, K. (1999). Newton: Pinball wizard? *Popular Measuremen*t, 5-7.

Ludlow, L. & Haley, S. (1995). Rasch model logits: Interpretation, use, and transformation. *Educational and Psychological Measurement, 55*(6), 967-975.

Ludlow, L. (1985). A strategy for the graphical representation of Rasch model residuals. *Educational and Psychological Measurement, 45,* 851-859.

Ludlow, L. (1983). The analysis of Rasch model residuals. Unpublished doctoral dissertation. University of Chicago.

Lundberg, C. & Schreiner, L. (2004). Quality and frequency of faculty-student interaction as predictors of learning: An analysis by race/ethnicity. *Journal of College Student Development, 45*(5), 549-565.

Mansson, D., Myers, S., & Martin, M. (2012). Students' communicative attributes and their out-of-class communication with instructors. *Atlantic Journal of Communication, 20*(4), 237-247.

Mara, M. & Mara, A. (2011). Finding and analytic frame for faculty-student interaction within faculty-in-residence programs. *Innovation in Higher Education, 36*, 71-82.

Martin, M., & Myers, S. (2006). Students' communication traits and their out-of-class communication with their instructors. *Communication Research Reports, 23*(4), 283-289.

Mayhew, M., Rockenbach, A., Bowman, N., Seifert, T., Wolniak, G., Pascarella, E., Terenzini, P. (2016). *How college affects students: 21st century evidence that higher education works, volume 3*. Jossey-Bass.

Micari, M. & Pazos, P. (2012). Connecting to the professor: Impact of the student-faculty relationship in a highly challenging course. *College Teaching, 60*(2), 41-47.

Miller-Ott, A. (2016). Helicopter parenting, family communication patterns, and out-of-class communication with college instructors. *Communication Research Reports, 33*(2), 173-176.

Myers, S., Martin, M., & Knapp, J., (2005). Perceived instructor communicative behaviors as a predictor of student participation in out of class communication. *Communication Quarterly, 53*(4), 437-450.

Nadler, K., & Nadler, J. (2001). The roles of sex, empathy, and credibility in out-of-class communication between faculty and students. *Women's Studies in Communication, 24*(2), 241-261.

Nadler, M. & Nadler, L. (2000). Our of class communication between faculty and students: A faculty perspective. *Communication Studies, 51*(2), 176-188.

Pascarella, E. (1985). College environmental influences on learning and cognitive development: A critical review and synthesis. In J. Smart (Ed.). *Higher education: Handbook of theory and research.* Agathon Press, Inc.

Pascarella, E. (1980). Student-faculty informal contact and college outcomes. *Review of Educational Research, 50*(4), 545-595.

Pascarella, E. & Terenzini, P. (1978). Student-faculty informal relationships and freshman year educational outcomes. *The Journal of Educational Research, 71*(4), 183-189.

Pascarella, E. & Terenzini, P. (1977). Patterns of student-faculty informal interaction beyond the classroom and voluntary freshman attrition. *The Journal of Higher Education, 48*(5), 540-552.

Pascarella, E., Terenzini, P., Hibel, J. (1978). Student-faculty interactional settings and their relationship to predicted academic performance. *The Journal of Higher Education, 4*(5), 450-463.

Rasch, G. (1960/1980). *Probabilistic models for some intelligences and attainment tests.* The University of Chicago Press.

Rasch G. (1966). An individualistic approach to item analysis. In P. F. Lazarsfeld & N. W. Henry (Eds.), *Readings in mathematical social science,* (pp. 84-108). Science Research Associates.

Reynolds, K. & Ludlow, L. (in press). Does instructor availability matter? Evidence from course evaluations. *Journal on Excellence in College Teaching.*

Rosenthal, G., Folse, E., Alleman, N., Boudreaux, D., Soper, B., & Von Bergen, C. (2000). The one-to-one survey: traditional versus non-traditional student satisfaction with professors during one-to-one contacts. *College Student Journal, 34*(2), 315-320.

Rowan-Kenyon, H., Martínez Alemán, A., & Savitz-Romer, M. (2018). *Technology and engagement: Making technology work for first-generation college students*. Rutgers University Press.

Sarapin, S. H., & Morris, P. L. (2015). Faculty and facebook friending: Instructor–student online social communication from the professor's perspective. *The Internet and Higher Education, 27*, 14-23.

Sax, L., Bryant, A., & Harper, C. (2005). The differential effects of student-faculty interaction on college outcomes for men and women. *Journal of College Student Development, 46*(6), 642-657.

Sheer, V. Fung, T. (2007). Can email communication enhance professor-student relationship and student evaluation of professor?: Some empirical evidence. *Journal of Educational Computing Research, 37*(3), 289-306.

Shye, S. (2015). New directions in facet theory. In *15th International Facet Theory Conference Proceedings* (pp. 147-162). New York City, NY.

Shye, S. (1998). Modern facet theory: Content design and measurement in behavioral research. *European Journal of Psychological Assessment, 14*(2), 160-171.

Sidelinger, R., Blen, D., McMullen, A. & Nyeste, M. (2015). Academic and social integration in the basic communication course: predictors of students' out-of-class communication and academic learning. *Communication Studies, 66*(1), 63-84.

Sinclaire, J. (2014). An empirical investigation of student satisfaction with college courses. *Research in Higher Education Journal, 22.*

Smith, R. & Miao, C. (1994). Assessing unidimensionality for Rasch measurement. In M. Wilson, (Ed.), *Objective measurement theory and practice volume 2* (pp. 316-327). Ablex Publishing Company.

Smith, R. & Plackner, C. (2009). The family approach to assessing fit in Rasch measurement. *Journal of Applied Measurement, 10*(4), 424-437.

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). *Rasch fit statistics and sample size considerations for polytomous data.* BMC Medical Research Methodology, 8(33).

Snow, S. (1973). Correlates of faculty-student interaction. *Sociology of Education, 46*(4), 489-498.

Solorzano, D., Ceja, M. & Yosso, T. (2000). Critical race theory, racial microaggressions, and campus climate: The experiences of African American college students. *The Journal of Negro Education, 69*, 60-73.

Stephens, K., Houser, M., & Cowan, R. (2009). R U able to meat me: The impact of students' overly casual email messages to instructors. *Communication Education, 58*(3), 303-326.

Tennant, A. & Pallant, J. (2006). Unidimensionality matters! (A tale of two smiths?). *Rasch Measurement Transactions, 20*(1), 1048-51.

Terenzini, P., Pascarella, E., & Blimling, G. (1996). Students' out-of-class experiences and their influence on learning and cognitive development: A literature review. *Journal of College Student Development, 40*(5), 619-623.

Thiele, M. (2016). Resource or obstacle? Classed reports of student-faculty relations. *The Sociological Quarterly, 57,* 333-355.

Theophilides, C. & Terenzini, P. (1981). The relation between nonclassroom contact with faculty and students' perceptions of instructional quality. *Research in Higher Education, 15*(3), 255-269.

Thurstone L. (1959). Attitudes can be measured. In Thurstone L.L. *The measurement of values*. The University of Chicago Press.

Trolian, T. & Parker, E. (2017). Moderating influences of student-faculty interactions on students' graduate and professional school aspirations. *Journal of College Student Development, 58*(8), 1261-1267.

Umbach, P. & Wawrzynski, M. (2005). Faculty do matter: The role of college faculty in student learning and engagement. *Research in Higher Education, 6*(2), 153-184.

Welfare, L. E. & Sackett, C. R. (2011). The authorship determination process in student-faculty collaborative research. *Journal of Counseling & Development, 89*(4), 479-487.

Wilson, R., Woods, L., & Gaff, J. (1974). Social-psychological accessibility and faculty-student interaction beyond the classroom. *Sociology of Education, 47(*1). 74-92.

Wolfe, E. W., & Smith, E. V. (2007). Instrument development tools and activities for measure validation using Rasch models: Part II—Validation activities. *Journal of Applied Measurement, 8*(2), 204-234.

Wright, B. A. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14*(2), 97-116.

Wright B. D. (1967). Sample-free test calibration and person measurement. In *Proceedings*

of the 1967 Invitational Conference on Testing Problems. Princeton: Educational

Testing Service, 85-101.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis.* MESA Press.

Zhao, Q., Ahn, S., Meyers, R., Timmerman, C., & Fonner, K. (2012). Exploring students' use of email for out-of-class communication: Frequency, satisfaction, and learning self-efficacy. *Journal on Excellence in College Teaching, 23*(4), 5-32.

Zhang, Q. (2006). Immediacy and out-of-class communication: A cross-cultural comparison. *International Journal of Intercultural Relations, 30*, 33-50.

Appendix A: Hypothesized Continua

Table A.1

Physical Accessibility

	Arranged Meetings	Informal Chance Encounters	Email
XX's physical accessibility...	Is *always* willing to accommodate students' schedules and flexible in scheduling meetings Is *generally* willing to accommodate students' schedules and flexible in scheduling meetings Is *sometimes* willing to accommodate students' schedules and flexible in scheduling meetings Is *rarely* willing to accommodate students' schedules and flexible in scheduling meetings Is *not* willing to accommodate students' schedules and flexible in scheduling meetings	*Constantly* has chance informal encounters with students *Frequently* has chance informal encounters with students *Sometimes* has chance informal encounters with students *Never* has chance informal encounters with students	*Always* responds to emails within the time frame needed *Usually* responds to emails within the time frame needed *Sometimes* responds to emails within the time frame needed *Rarely* responds to emails within the time frame needed *Never* responds to emails within the time frame needed

Table A.2

Social Engagement

	Arranged Meetings	**Informal Chance Encounters**	**Email**
XX's social engagement...	Is *completely engaged and not at all distracted* during meetings with students in the office Is *generally interested and not distracted* during meetings with students in the office Is *somewhat distracted* during meetings with students in the office *Does not engage meaningfully* with students during meetings in the office	Seems *excited and enthusiastic* to have chance encounters with students Seems *happy* to have chance encounters with students *Does not really react* to chance encounters with students Seems *reluctant* to have chance encounters with students Seems *annoyed* during chance encounters with students	Provides *thoughtful responses thoroughly addressing students' points and questions* Provides *adequate responses addressing students' points and questions* Provides *responses that do not address students' points and questions*

Appendix B: Narrative Descriptions of Faculty at High, Medium, and Low Levels of Physical Accessibility and Social Engagement

Table B.1

Physical Accessibility Narratives

Facet	Level	Description
Arranged Meetings	High	A faculty member who is high on the physical accessibility domain of willingness to engage in OCC provides many opportunities for students to meet with him/her in the office. This might include keeping regularly scheduled office hours week to week that are sufficient to accommodate student schedules. It might also include a willingness to meet outside of regularly scheduled office hours if necessary. A faculty member may still be high on this dimension without have regular weekly office hours; in this case, the faculty member is very flexible when it comes to scheduling office meetings with students. Along with flexibility, a faculty member who is high on physical accessibility for office meetings also meets with students within a reasonable time frame when meetings are requested. A faculty member does not have to be constantly available to fall into the high category; it is acknowledged that every individual will sometimes have unforeseeable conflicts or occasional travel commitments; however, these should not generally interfere with availability for meeting with students.
	Medium	A faculty member who is medium on the physical accessibility domain of willingness to engage in OCC sometimes provides adequate opportunities for students to meet with him/her in the office. This might include keeping regularly scheduled office hours, although these hours might not be sufficient to accommodate student needs. It might also include a willingness to meet outside of office hours, but only in exceptional circumstances. If a faculty member does not keep regular office hours, they may fall into the medium category by sometimes accommodating student requests, but being consistently unavailable at specific times (for example, certain days of the week). A faculty member who is medium on physical accessibility generally meets with students within a reasonable time frame when meetings are requested, but occasionally does not accommodate requests within a reasonable time frame. A faculty member in the medium category may travel frequently enough that is can be difficult to arrange meetings.
	Low	A faculty member who is low on the physical accessibility domain of willingness to engage in OCC does not provide adequate opportunities for students to meet with him/her in the office. This might include frequently canceling scheduled office hours, with or without notice. It might also include being unwilling to meet with students outside of office hours. If regular office hours are not kept, a faculty member who

		is low on physical accessibility is very difficult to schedule meetings with and often does not accommodate requests with meetings within a reason time frame. A faculty member in the low category may travel frequently, or only be on campus during class time.
Informal Chance Encounters	High	A faculty member who is high on the physical accessibility domain is frequently on campus, allowing for many opportunities to have chance encounters with students. S/he is a visible presence around campus (or at least around the building housing her/his office), and does not spend all of her/his on-campus time shut in the office. A faculty member in the high category is present on campus even on days where s/he does not have formal commitments or reasons for being there (e.g., teaching class, scheduled office hours, meetings).
	Medium	A faculty member who is medium on the physical accessibility domain is generally on campus, allowing for some opportunities to have chance encounters with students. S/he is sometimes seen around campus (or the building housing his/her office), but is not necessarily a visible presence and may spend most of her/his on-campus time shut in the office. A faculty member in the medium category may occasionally be present on campus on days where s/he does not have formal commitments or reasons for being there (e.g., teaching class, scheduled office hours, meetings).
	Low	A faculty member who is low on the physical accessibility domain is not generally present on campus and does not have many opportunities to have chance encounters with students. S/he is never seen around campus (or the building housing her/his office) and spends all of her/his on-campus time shut in the office. A faculty member in the low category is never present on campus on days where s/he does not have formal commitments or reasons for being there (e.g., teaching class, scheduled office hours, meetings).
Email	High	A faculty member who is high on the physical accessibility dimension always responds to emails within the time that is needed (recognizing that "time needed" may vary based on circumstance). Students do not feel as though they are unable to move forward with classwork (or other relevant work) because they are awaiting email responses. Even if s/he is unable to provide a full response immediately, a faculty member in the high category will at least send a brief note letting students know when they might expect a more detailed response. If a faculty member in the low category receives an email about something they perceive to be unimportant or a non-issue, they will still always respond.
	Medium	A faculty member who is medium on the physical accessibility dimension generally responds to emails within the time frame that is needed (recognizing that "time needed" may vary based on circumstance). Students occasionally feels as though they are unable to move forward with classwork (or other relevant

	work) because they are awaiting email responses, but this is not a frequent occurrence. If s/he cannot provide a full response immediately, a faculty member in the medium category may let a student know, but does not provide an indication of when the student might expect a more detailed response. If a faculty member in the low category receives an email about something they perceive to be unimportant or a non-issue, they will still generally respond.
Low	A faculty member who is low on the physical accessibility dimension does not respond to emails within the time frame that is needed (recognizing that "time needed" may vary based on circumstance). Students frequently feel as though they are unable to move forward with classwork (or other relevant work) because they are awaiting email responses. If s/he cannot provide a full response immediately, a faculty member in the low category will not respond at all. If a faculty member in the low category receives an email about something they perceive to be unimportant or a non-issue, they will not respond.

Table B.2

Social Engagement Narratives

Facet	Level	Description
Arranged Meetings	High	A faculty member who is high on the social accessibility dimension is completely engaged during office meetings with students. S/he carefully listens to student concerns, responds thoughtfully, and does not seem at all distracted by other matters. Although meetings may have scheduled endpoints, faculty members in the high category do not rush students out the door and generally do not end meeting until all student concerns have been addressed. If interruptions, such as phone calls or knocks on the door occur, the faculty member prioritizes the student meeting (emergencies excepted).
	Medium	A faculty member who is medium on the social accessibility dimension is generally engaged during office meetings with students. S/he listens to student concerns and provides responses that are generally adequate. S/he may seem somewhat impatient or distracted during meetings with students. S/he may end meeting before all student concerns are adequately resolved. If interruptions, such as phone calls or knocks on the door occur, the faculty member may ask students to wait a few minutes while the matter is dealt with.
	Low	A faculty member who is low on the social accessibility dimension does not seem engaged during office meetings with students. S/he does not listen carefully to student concerns, provides inadequate responses, and frequently seems distracted. Faculty in the low category are always trying to get students out the door as quickly as possible. If interruptions, such as phone calls or knocks on the door occur, the faculty member does not prioritize the student meeting and may end the student meeting early to deal with other matters.
Informal Chance Encounters	High	A faculty member who is high on the social accessibility domain seems happy and enthusiastic to have chance encounters with students. Although s/he may not always have time to stop for a full conversation, s/he always greets students when s/he sees them. Time permitting, a faculty member in the high category does stop to have chats with students around campus or in the halls, whether it be small talk or academic matters. The faculty member recognizes individual students and seems genuinely pleased to see students outside of class.
	Medium	A faculty member who is medium on the social accessibility domain seems indifferent to having chance encounters with students. S/he generally does not stop to have full conversations with students around campus or in the halls, but may wave hello and offer a greeting. A faculty member in the medium

		category may not always recognize individual students and does not seem to have strong feelings about seeing students outside of class.
	Low	A faculty member who is low on the social accessibility domain does not seem to enjoy having chance encounters with students. S/he never stops to have full conversations with students around campus or in the halls and frequently does not even say hello or offer a greeting. A faculty member in the low category does not recognize individual students and seems to dislike seeing students outside of class.
Email	High	A faculty member who is high on the social accessibility dimension always provides email responses that adequately address student questions or concerns. (This does not mean that responses must be lengthy.) If the faculty member is unclear about a student question/concern, s/he follows up with clarifying questions. A faculty member in the high category invites students to follow up with her/him about any additional issues or concerns. If the answer to a student question is contained in an easily accessible document, such as the course syllabus, the faculty member encourages the student to consult the appropriate document and does not come off as annoyed while doing so.
	Medium	A faculty member who is medium on the social accessibility dimension sometimes provides email responses that adequately address student questions or concerns. (This does not mean that response must be lengthy.) If the faculty member is unclear about a student question/concern, s/he will say so, but will not follow up with clarifying questions. If the answer to a student question is contained in an easily accessible document, the faculty member encourages the student to consult the appropriate document but may come off as annoyed while doing so.
	Low	A faculty member who is low on the social accessibility dimension does not provide email responses that adequately address student questions or concerns. If a faculty member is unclear about a student/question or concern, s/he will not address it. If the answer to a student question is contained in an easily accessible document, the faculty member says so and is clearly annoyed that the question was asked via email.

Appendix C: Brief Descriptors of High and Low Levels

Table C.1

Continua for the Facetized Constructs of Physical Accessibility and Social Engagement

Dimension	Facet M: Arranged Meetings	Facet I: Informal Change Encounters	Facet E: Email
Physical Accessibility	sufficient, flexible meeting times (high) insufficient, rigid meeting times (low)	a constant presence around campus (high) never seen around campus (low)	timely and consistent email responses (high) slow and inconsistent email responses (low)
Social Engagement	attentive, engaged during meetings (high) distracted, indifferent during meetings (low)	happy to see students around campus (high) avoids acknowledging students around campus (low)	helpful email responses that are professional in tone (high) unhelpful email responses that are rude in tone (low)

Appendix D: Facet Level Combinations for Each Scale

Table D.1

Initial Physical Accessibility Scale Combinations (Pre-pilot and Pilot)

Item	Facet M: Arranged Meetings	Facet C: Chance Encounters	Facet E: Email
1	High (3)	High (3)	High (3)
2	High (3)	High (3)	Medium (2)
3	High (3)	Medium (2)	High (3)
4	Medium (2)	Medium (2)	High (3)
5	Medium (2)	Medium (2)	Medium (2)
6	Medium (2)	Low (1)	Medium (2)
7	Medium (2)	Medium (2)	Low (1)
8	Low (1)	Low (1)	Medium (2)
9	Low (1)	Low (1)	Low (1)

Table D.2

Initial Social Engagement Scale Combinations (Pre-pilot and Pilot)

Item	Facet M: Arranged Meetings	Facet C: Chance Encounters	Facet E: Email
1	High (3)	High (3)	High (3)
2	High (3)	High (3)	Medium (2)
3	High (3)	Medium (2)	High (3)
4	Medium (2)	Medium (2)	High (3)
5	Medium (2)	Medium (2)	Medium (2)
6	Medium (2)	Low (1)	Medium (2)
7	Medium (2)	Medium (2)	Low (1)
8	Low (1)	Low (1)	Medium (2)

9	Low (1)	Low (1)	Low (1)

Table D.3

Final Physical Accessibility Scale Combinations

Item	Facet M: Arranged Meetings	Facet C: Chance Encounters	Facet E: Email
1	High (3)	High (3)	High (3)
2	High (3)	High (3)	Medium (2)
3	Medium (2)	Medium (2)	High (3)
4	Medium (2)	Medium (2)	Medium (2)
5	Medium (2)	Medium (2)	Low (1)
6	Low (1)	Low (1)	Medium (2)
7	Low (1)	Low (1)	Low (1)

Table D.4

Final Social Engagement Scale Combinations

Item	Facet M: Arranged Meetings	Facet C: Chance Encounters	Facet E: Email
1	High (3)	High (3)	High (3)
2	High (3)	High (3)	Medium (2)
3	Medium (2)	Medium (2)	High (3)
4	Medium (2)	Medium (2)	Medium (2)
5	Medium (2)	Medium (2)	Low (1)
6	Low (1)	Low (1)	Medium (2)
7	Low (1)	Low (1)	Low (1)

Appendix E: First Drafts and Final Items

Table E.1

Physical Accessibility Scale Item First Drafts and Final Items

Item First Draft	Final Item
M3, C3, E3 (9) Professor X keeps regular office hours this week and is also accommodating to student requests for meetings outside of these times. S/he is a constant presence on campus and students often see him/her in the halls of the building where her/his office is located. Professor X always responds to emails in a timely manner and students never feel as though they are kept from moving forward on assignments while awaiting responses. How does [professor] compare to Professor X? • S/he is much more accessible than Professor X. • S/he is more accessible than Professor X. • S/he is about as accessible as Professor X. • S/he is less accessible than Professor X. • S/he is much less accessible than Professor X.	M3, C3, E3 (9) Professor C holds regular office hours and meets students outside of these times when necessary. They seem to always be on campus and are often seen by students. Students can count on Professor C to respond to emails in a timely manner. How does the faculty member you have in mind compare to Professor C? • They are much more accessible than Professor C. • They are a little more accessible than Professor C. • They and Professor C are about the same. • Professor C is a little more accessible than them. • Professor C is much more accessible than them.
M3, C3, E2 (8) Professor X does not keep weekly office hours, but is always willing to accommodate student meetings either in-person or via Skype. S/he is on campus every day of the week and frequently crosses paths with students. Professor X does not consistently respond to emails in a timely manner, sometimes creating stressful situations for students.	M3, C3, E2 (8) Professor A does not hold regular weekly office hours, but often meets with students either in-person or through videochat. They are on campus every day and students frequently see Professor A outside of class. However, Professor A is inconsistent in responding to emails in a timely manner. How does the faculty member you have in mind compare to Professor A?

How does [professor] compare to Professor X? • S/he is much more accessible than Professor X. • S/he is more accessible than Professor X. • S/he is about as accessible as Professor X. • S/he is less accessible than Professor X. • S/he is much less accessible than Professor X.	• They are much more accessible than Professor A. • They are a little more accessible than Professor A. • They and Professor A are about the same. • Professor A is a little more accessible than them. • Professor A is much more accessible than them.
M3, C2, E3 (8) Professor X keeps regular office hours each day of the week and meets with students via Skype if necessary. S/he is usually on campus, but is not a particularly visible presence to students and spends most of her/his time in her/his office. Professor X responds quickly to emails; students know they will usually receive a response within 24 hours. How does [professor] compare to Professor X? • S/he is much more accessible than Professor X. • S/he is more accessible than Professor X. • S/he is about as accessible as Professor X. • S/he is less accessible than Professor X. • S/he is much less accessible than Professor X.	M3, C2, E3 (8) Not included in the final scale.
M2, C2, E3 (7) Professor X keeps scheduled office hours two days a week, but does not accommodate student requests for meetings outside of these times. S/he is sometimes seen around campus by students, but is not a frequent presence. Professor X always responds to student emails within one business day.	M2, C2, E3 (7) Professor H holds office hours twice a week and will only meet with students during these times. Students only see them outside of class occasionally, but Professor H usually responds to emails within one weekday.

How does [professor] compare to Professor X? • S/he is much more accessible than Professor X. • S/he is more accessible than Professor X. • S/he is about as accessible as Professor X. • S/he is less accessible than Professor X. • S/he is much less accessible than Professor X.	How does the faculty member you have in mind compare to Professor H? • They are much more accessible than Professor H. • They are a little more accessible than Professor H. • They and Professor H are about the same. • Professor H is a little more accessible than them. • Professor H is much more accessible than them.
M2, C2, E2 (6) Professor X keeps regular office hours, but they are not always sufficient; students sometimes must wait over a week for a meeting. S/he sometimes crosses paths with students around campus, but not on a regular basis. Professor X usually responds to emails in a timely manner, but sometimes students await a response for several business days. How does [professor] compare to Professor X? • S/he is much more accessible than Professor X. • S/he is more accessible than Professor X. • S/he is about as accessible as Professor X. • S/he is less accessible than Professor X. • S/he is much less accessible than Professor X.	M2, C2, E2 (6) Professor M cannot always quickly accommodate meeting requests they receive; students sometimes must wait over a week. They sometimes cross paths with students around campus and usually respond to emails in a timely manner, but students sometimes do not receive a reply for several days. How does the faculty member you have in mind compare to Professor M? • They are much more accessible than Professor M. • They are a little more accessible than Professor M. • They and Professor M are about the same. • Professor M is a little more accessible than them. • Professor M is much more accessible than them.
M2, C1, E2 (5) Professor X has scheduled office hours, but sometimes is not available during those times. S/he is never seen around campus outside of her/his office or scheduled class times. Professor X sometimes responds to emails within 24	M2, C1, E2 (5) Not included in the final scale.

hours; other times students may await responses for over a week. How does [professor] compare to Professor X? • S/he is much more accessible than Professor X. • S/he is more accessible than Professor X. • S/he is about as accessible as Professor X. • S/he is less accessible than Professor X. • S/he is much less accessible than Professor X.	
M2, C2, E1 (5) Professor X has scheduled office hours two days per week, but sometimes cancels them without warning. S/he is occasionally seen in outside of his/her office by students. Professor X rarely responds to emails in a timely manner; students who pose questions via email often find they go unanswered. How does [professor] compare to Professor X? • S/he is much more accessible than Professor X. • S/he is more accessible than Professor X. • S/he is about as accessible as Professor X. • S/he is less accessible than Professor X. • S/he is much less accessible than Professor X.	M2, C2, E1 (5) Professor O has scheduled office hours multiple days throughout the week, but sometimes cancels them at the last minute. Students occasionally see them around campus outside of class, but Professor O rarely responds to emails in a timely manner; students often find emails go unanswered. How does the faculty member you have in mind compare to Professor O? • They are much more accessible than Professor O. • They are a little more accessible than Professor O. • They and Professor O are about the same. • Professor O is a little more accessible than them. • Professor O is much more accessible than them.
M1, C1, E2 (4) Professor X does not keep scheduled office hours and typically does not accommodate student requests for meetings. S/he is never seen by students outside of class times. Professor X usually responds to student emails; however,	M1, C1, E2 (4) Professor J does not have scheduled office hours and typically does not accommodate student requests for meetings; students rarely see them outside of class at all.

<table>
<tr>
<td>students sometimes must wait over a week for a response.

How does [professor] compare to Professor X?
S/he is much more accessible than Professor X.
S/he is more accessible than Professor X.
S/he is about as accessible as Professor X.
S/he is less accessible than Professor X.
S/he is much less accessible than Professor X.</td>
<td>Professor J responds to student emails, but not always in a timely manner.

How does the faculty member you have in mind compare to Professor J?
They are much more accessible than Professor J.
They are a little more accessible than Professor J.
They and Professor J are about the same.
Professor J is a little more accessible than them.
Professor J is much more accessible than them.</td>
</tr>
<tr>
<td>M1, C1, E1 (3)

Professor X is almost never available to meet with students; s/he keeps one office hour a week and does not accommodate student requests to meet at other times. S/he is never seen around campus or in the halls of the building where his/her office is located; students only see him/her during class. Professor X rarely responds to emails in a timely manner; it is not uncommon for students to receive no response at all.

How does [professor] compare to Professor X?
S/he is much more accessible than Professor X.
S/he is more accessible than Professor X.
S/he is about as accessible as Professor X.
S/he is less accessible than Professor X.
S/he is much less accessible than Professor X.</td>
<td>M1, C1, E1 (3)

Professor B is rarely available to meet with students; they keep few office hours and do not accommodate requests to meet at other times. The only time students see Professor B is during class. Professor B rarely responds to emails in a timely manner and sometimes does not respond at all.

How does the faculty member you have in mind compare to Professor B?
They are much more accessible than Professor B.
They are a little more accessible than Professor B.
They and Professor B are about the same.
Professor B is a little more accessible than them.
Professor B is much more accessible than them.</td>
</tr>
</table>

Table E.2

Social Engagement Scale Item First Drafts and Final Items

<table>
<tr><th>Item First Draft</th><th>Final Item</th></tr>
<tr><td>M3, C3, E3 (9)

Professor X is clearly engaged during meetings with students; s/he listens carefully to student concerns or questions and engages in thoughtful discussion. When s/he encounters students around campus, Professor X gives enthusiastic greetings and seems happy to see students. Professor X's responses to student emails are always courteous and respectful.

How does [professor] compare to Professor X?
S/he is much more accessible than Professor X.
S/he is more accessible than Professor X.
S/he is about as accessible as Professor X.
S/he is less accessible than Professor X.
S/he is much less accessible than Professor X.</td><td>M3, C3, E3 (9)

Professor K is clearly engaged during meetings with students; they listen to students carefully, ask questions, and engage in thoughtful discussion. They seem happy to see students around campus and often stop to chat. Professor K's responses to emails are always courteous and clearly address students' questions or concerns.

How does the faculty member you have in mind compare to Professor K?
They are much more engaged than Professor K.
They are a little more engaged than Professor K.
They are Professor K are about the same.
Professor K is a little more engaged than them.
Professor K is much more engaged than them.</td></tr>
<tr><td>M3, C3, E2 (8)

Professor X is fully present during meetings with students; s/he does not seem at all distracted by other matters. When students see Professor X in the halls, s/he always offers a friendly greeting. Professor X's responses to emails are sometimes terse, but never rude or disrespectful.

How does [professor] compare to Professor X?
S/he is much more accessible than Professor X.</td><td>M3, C3, E2 (8)

Professor L clearly prioritizes students during meetings and does not seem at all distracted by other matters. They often greet students they see around campus. Professor L's responses to emails are sometimes abrupt, but never rude or disrespectful and usually adequately address students' concerns or questions.

How does the faculty member you have in mind compare to Professor L?
They are much more engaged than Professor L.</td></tr>
</table>

• S/he is more accessible than Professor X. • S/he is about as accessible as Professor X. • S/he is less accessible than Professor X. • S/he is much less accessible than Professor X.	• They are a little more engaged than Professor L. • They are Professor L are about the same. • Professor L is a little more engaged than them. • Professor L is much more engaged than them.
M3, C2, E3 (8) Professor X clearly prioritizes students during meetings; if disruptions arise, s/he finishes the student meeting first, without making students feel as though they are rushed out the door. Professor X sometimes waves to students when they pass on campus, but other times does not seem to recognize them. Professor X's responses to student emails always address student concerns in a professional tone. How does [professor] compare to Professor X? • S/he is much more accessible than Professor X. • S/he is more accessible than Professor X. • S/he is about as accessible as Professor X. • S/he is less accessible than Professor X. • S/he is much less accessible than Professor X.	M3, C2, E3 (8) Not included in the final scale.
M2, C2, E3 (7) Professor X seems to listen to students' concerns and questions during meetings, but does not always engage in thoughtful discussion and may provide generic comments. S/he will sometimes greet students when they cross paths around campus, but does not do so enthusiastically. Professor X's responses	M2, C2, E3 (7) Professor T listens to students' concerns and questions during meetings, but does not always engage in thoughtful conversation. They acknowledge students inconsistently when crossing paths around campus. Professor T's responses to students' emails, however, are thorough and respectful.

to students' emails are thorough and cordial. How does [professor] compare to Professor X? • S/he is much more accessible than Professor X. • S/he is more accessible than Professor X. • S/he is about as accessible as Professor X. • S/he is less accessible than Professor X. • S/he is much less accessible than Professor X.	How does the faculty member you have in mind compare to Professor T? • They are much more engaged than Professor T. • They are a little more engaged than Professor T. • They are Professor T are about the same. • Professor T is a little more engaged than them. • Professor T is much more engaged than them.
M2, C2, E2 (6) Professor X sometimes seems distracted during meetings with students, occasionally making students feel as though they are being rushed out the door. Professor X may smile at students when s/he encounters them around campus, but does not offer a verbal greeting. Professor X's responses to students' emails are always courteous, although s/he does not always clearly and thoughtfully address students' questions and concerns. How does [professor] compare to Professor X? • S/he is much more accessible than Professor X. • S/he is more accessible than Professor X. • S/he is about as accessible as Professor X. • S/he is less accessible than Professor X. • S/he is much less accessible than Professor X.	M2, C2, E2 (6) Professor Y occasionally seems distracted during meetings with students. They may acknowledge students around campus, but rarely start up a conversation. Professor Y's responses to emails are courteous, although they do not always clearly address students' questions and concerns. How does the faculty member you have in mind compare to Professor Y? • They are much more engaged than Professor Y. • They are a little more engaged than Professor Y. • They are Professor Y are about the same. • Professor Y is a little more engaged than them. • Professor Y is much more engaged than them.
M2, C1, E2 (5) Professor X is generally focused during meetings with students, although s/he may	M2, C1, E2 (5) Not included in the final scale.

not always engage in thoughtful discussions with students. Professor X rarely acknowledges students when s/he encounters them outside of class or office hours. Professor X's responses to student emails may be generic, but are generally courteous. How does [professor] compare to Professor X? • S/he is much more accessible than Professor X. • S/he is more accessible than Professor X. • S/he is about as accessible as Professor X. • S/he is less accessible than Professor X. • S/he is much less accessible than Professor X.	
M2, C2, E1 (5) Professor X is not always fully engaged during meetings with students; her/his attention sometimes appears to be elsewhere. When s/he sees students around campus, Professor X occasionally offers a wave of recognition. Professor X's responses to student emails do not clearly address students' concerns and may be rude in tone. How does [professor] compare to Professor X? • S/he is much more accessible than Professor X. • S/he is more accessible than Professor X. • S/he is about as accessible as Professor X. • S/he is less accessible than Professor X. • S/he is much less accessible than Professor X.	M2, C2, E1 (5) Professor S is not always fully engaged during meetings with students; their attention sometimes appears to be elsewhere. When they see students around campus, Professor S occasionally offers a wave of recognition, but does not go out of their way to engage in conversation. Professor S's responses to emails are often unclear and may appear rude in tone. How does the faculty member you have in mind compare to Professor S? • They are much more engaged than Professor S. • They are a little more engaged than Professor S. • They are Professor S are about the same. • Professor S is a little more enagaged than them. • Professor S is much more engaged than them.
M1, C1, E2 (4)	M1, C1, E2 (4)

<table>
<tr>
<td>Professor X is clearly distracted during meetings with students; s/he routinely appears to be occupied with other tasks, such as answering emails. S/he does not go out of her/his way to greet students and seemed annoyed if students initiate conversation. Professor X’s responses to student emails are generally clear and professional.

How does [professor] compare to Professor X?
• S/he is much more accessible than Professor X.
• S/he is more accessible than Professor X.
• S/he is about as accessible as Professor X.
• S/he is less accessible than Professor X.
• S/he is much less accessible than Professor X.</td>
<td>Professor G is often preoccupied with other tasks during meetings with students. They do not go out of their way to greet students around campus and seem annoyed if students try to initiate conversation. However, Professor G’s responses to emails are generally clear and professional.

How does the faculty member you have in mind compare to Professor G?
• They are much more engaged than Professor G.
• They are a little more engaged than Professor G.
• They are Professor G are about the same.
• Professor G is a little more engaged than them.
• Professor G is much more engaged than them.</td>
</tr>
<tr>
<td>M1, C1, E1 (3)

Professor X seems annoyed during meetings with students and would clearly rather spend time on other activities. S/he does not acknowledge students when s/he sees them around campus. Professor X’s responses to students are disrespectful in tone, and do not clearly address student concerns or questions.

How does [professor] compare to Professor X?
• S/he is much more accessible than Professor X.
• S/he is more accessible than Professor X.
• S/he is about as accessible as Professor X.
• S/he is less accessible than Professor X.
• S/he is much less accessible than Professor X.</td>
<td>M1, C1, E1 (3)

Professor R often seems annoyed during meetings with students and would clearly rather spend time on other activities. They do not acknowledge students around campus and their responses to emails are often disrespectful in tone, and do not clearly address student concerns or questions.

How does the faculty member you have in mind compare to Professor R?
• They are much more engaged than Professor R.
• They are a little more engaged than Professor R.
• They are Professor R are about the same.
• Professor R is a little more engaged than them.
• Professor R is much more engaged than them.</td>
</tr>
</table>

Appendix F: Focus Group Protocol

Opening script:

Thank you so much for being here today. The purpose of this focus group is to have a discussion about your thoughts on student-faculty interaction outside of class. Specifically, we are interested in exploring students' perceptions that faculty are available for interaction outside of the classroom and building a survey that can measure these perceptions. We really appreciate you taking time of your schedules to be a part of this—your thoughts on this important topic are extremely helpful in refining our research.

Please note that everything that is said in this room stays in this room. While the thoughts you provide will be used to inform future research, your comments will never be presented in such a way that they are associated with your name or other identifying information. I ask that you also respect the privacy of others by not sharing their comments when you leave this space. I will also be recording our conversation today so that I can refer back to it later. The recording will not be shared with anyone else.

Does anyone have any questions before we begin?

Warm Up:

We're going to begin this conversation with a warm-up exercise related to student-faculty interaction outside of class. Specifically, let's think about how you know whether or not a faculty member is available for interaction outside of class. We're going to craft a narrative description as a group of what a "highly available" faculty member looks like. We can start by making a list and then will work to craft a description.

- What does a faculty member who is "highly available" outside of class look like?
- What kinds of actions do they take?
- What kinds of behaviors do they have?
- Is there anything that it is important they NOT do?

Now let's repeat this exercise for a faculty member who is NOT available outside of class. We can follow the same process—starting with bullet points and then writing a narrative description.

- What does a faculty member who is "not available" outside of class look like?
- What kinds of actions do they take?
- What kinds of behaviors do they have?
- Is there anything that it is important they NOT do?

Questions:

We are working on a research project to build a survey measuring students' perceptions that faculty are available outside of class. Now that we've built descriptions of available and unavailable faculty members as a group, let's spend some time talking about how this might translate to a survey.

Our current thinking on this topic is that faculty availability to students outside of class is defined by two parts: physical accessibility and social engagement. Physical accessibility refers to students' opportunities for interaction with the faculty member—for example, do they frequently occupy the same physical spaces outside of class? Social engagement refers to faculty members' behaviors when interacting with students outside of class—for example, do they seem to be listening to students and fully present in conversations? Does this seem like an adequate definition of faculty availability?

Probing questions:

- Is it appropriate to include both physical accessibility and social engagement in a survey about students' perceptions of faculty availability outside of class?
- Can you think of anything else you would add to this definition?
- How does this definition compare to the descriptions we generated during the warm-up exercise?
-

We also think that an instrument measuring students' perceptions of faculty availability outside of class should capture the different contexts in which this students and faculty interact outside of class. We think these spaces are arranged meetings (such as meeting during offices hours), informal interaction that occurs by chance (such as passing a faculty member in a hallway), and via email or other electronic communication. Do these seem adequate to capture the contexts in which students and faculty may interact outside of class?

Probing questions:

- Is it appropriate to include all of these contexts in a survey about students' perceptions of faculty availability outside of class?
- Are there any other contexts that you would add?

As a final activity, we have a draft of the survey we are working to develop. Please take a look at these survey questions, which were written based on the working definition of students' perceptions of faculty availability that we just discussed. There are questions specific to physical accessibility and questions specific to social engagement. Feel free to make any notes that you would like. Do these questions adequately reflect that definition?

Probing questions:

- Do these questions reflect the descriptions of faculty members with high and low availability that we created at the beginning of this session?
- Is there a clear progression of high availability and low availability represented for each of the two dimensions (accessibility and engagement)?
- Is the opening text of the survey clear? Do you understand how to respond to the questions?
- Are there any specific questions that have wording that is confusing?
- Do the response options for the questions make sense?

Closing:

We are reaching the end of our time. This discussion has been extremely helpful; thank you again for taking time out of your day to participate in this focus group. This is an ongoing research project; if you are interested in remaining involved or learning more about next steps, please feel free to keep in touch with me via email.

Appendix G: Final Administration Instrument

INSTRUCTIONS

This survey will ask you about your perceptions of faculty members' availability outside of class.

This survey requires you to keep in mind a specific faculty member **with whom you most recently had class** throughout all survey questions. Please respond to all questions with respect to **the faculty member with whom you most recently had class**. When responding, select the choice that best represents your typical experiences with that faculty member. All survey items use “they” or “them” as a gender-neutral, singular pronoun.

PAS ITEMS

Professor C holds regular office hours and meets students outside of these times when necessary. They seem to always be on campus and are often seen by students. Students can count on Professor C to respond to emails in a timely manner.
How does the faculty member you have in mind compare to Professor C?

- ◯ They are much more accessible than Professor C. (5)
- ◯ They are a little more accessible than Professor C. (4)
- ◯ They and Professor C are about the same. (3)
- ◯ Professor C is a little more accessible than them. (2)
- ◯ Professor C is much more accessible than them. (1)

Professor A does not hold regular weekly office hours, but often meets with students either in-person or through videochat. They are on campus every day and students frequently see Professor A outside of class. However, Professor A is inconsistent in responding to emails in a timely manner.
How does the faculty member you have in mind compare to Professor A?

- ◯ They are much more accessible than Professor A. (5)
- ◯ They are a little more accessible than Professor A. (4)
- ◯ They and Professor A are about the same. (3)
- ◯ Professor A is a little more accessible than them. (2)
- ◯ Professor A is much more accessible than them. (1)

Professor H holds office hours twice a week and will only meet with students during these times. Students only see them outside of class occasionally, but Professor H usually responds to emails within one weekday.
How does the faculty member you have in mind compare to Professor H?

- ○ They are much more accessible than Professor H. (5)
- ○ They are a little more accessible than Professor H. (4)
- ○ They and Professor H are about the same. (3)
- ○ Professor H is a little more accessible than them. (2)
- ○ Professor H is much more accessible than them. (1)

Professor M cannot always quickly accommodate meeting requests they receive; students sometimes must wait over a week. They sometimes cross paths with students around campus and usually respond to emails in a timely manner, but students sometimes do not receive a reply for several days.
How does the faculty member you have in mind compare to Professor M?

- ○ They are much more accessible than Professor M. (5)
- ○ They are a little more accessible than Professor M. (4)
- ○ They and Professor M are about the same. (3)
- ○ Professor M is a little more accessible than them. (2)
- ○ Professor M is much more accessible than them. (1)

Professor O has scheduled office hours multiple days throughout the week, but sometimes cancels them at the last minute. Students occasionally see them around campus outside of class, but Professor O rarely responds to emails in a timely manner; students often find emails go unanswered.
How does the faculty member you have in mind compare to Professor O?

- ○ They are much more accessible than Professor O. (5)
- ○ They are a little more accessible than Professor O. (4)
- ○ They and Professor O are about the same. (3)
- ○ Professor O is a little more accessible than them. (2)
- ○ Professor O is much more accessible than them. (1)

Professor J does not have scheduled office hours and typically does not accommodate student requests for meetings; students rarely see them outside of class at all. Professor J responds to student emails, but not always in a timely manner.
How does the faculty member you have in mind compare to Professor J?

- They are much more accessible than Professor J. (5)
- They are a little more accessible than Professor J. (4)
- They and Professor J are about the same. (3)
- Professor J is a little more accessible than them. (2)
- Professor J is much more accessible than them. (1)

Professor B is rarely available to meet with students; they keep few office hours and do not accommodate requests to meet at other times. The only time students see Professor B is during class. Professor B rarely responds to emails in a timely manner and sometimes does not respond at all.
How does the faculty member you have in mind compare to Professor B?

- They are much more accessible than Professor B. (5)
- They are a little more accessible than Professor B. (4)
- They and Professor B are about the same. (3)
- Professor B is a little more accessible than them. (2)
- Professor B is much more accessible than them. (1)

SES ITEMS

Professor K is clearly engaged during meetings with students; they listen to students carefully, ask questions, and engage in thoughtful discussion. They seem happy to see students around campus and often stop to chat. Professor K's responses to emails are always courteous and clearly address students' questions or concerns.
How does the faculty member you have in mind compare to Professor K?

- They are much more engaged than Professor K. (5)
- They are a little more engaged than Professor K. (4)
- They and Professor K are about the same. (3)
- Professor K is a little more engaged than them. (2)

- ○ Professor K is much more engaged than them. (1)

Professor L clearly prioritizes students during meetings and does not seem at all distracted by other matters. They often greet students they see around campus. Professor L's responses to emails are sometimes abrupt, but never rude or disrespectful and usually adequately address students' concerns or questions.
How does the faculty member you have in mind compare to Professor L?

- ○ They are much more engaged than Professor L. (5)
- ○ They are a little more engaged than Professor L. (4)
- ○ They and Professor L are about the same. (3)
- ○ Professor L is a little more engaged than them. (2)
- ○ Professor L is much more engaged than them. (1)

Professor T listens to students' concerns and questions during meetings, but does not always engage in thoughtful conversation. They acknowledge students inconsistently when crossing paths around campus. Professor T's responses to students' emails, however, are thorough and respectful.
How does the faculty member you have in mind compare to Professor T?

- ○ They are much more engaged than Professor T. (5)
- ○ They are a little more engaged than Professor T. (4)
- ○ They and Professor T are about the same. (3)
- ○ Professor T is a little more engaged than them. (2)
- ○ Professor T is much more engaged than them. (1)

Professor Y occasionally seems distracted during meetings with students. They may acknowledge students around campus, but rarely start up a conversation. Professor Y's responses to emails are courteous, although they do not always clearly address students' questions and concerns.
How does the faculty member you have in mind compare to Professor Y?

- ○ They are much more engaged than Professor Y. (5)
- ○ They are a little more engaged than Professor Y. (4)
- ○ They and Professor Y are about the same. (3)

- Professor Y is a little more engaged than them. (2)
- Professor Y is much more engaged than them. (1)

Professor S is not always fully engaged during meetings with students; their attention sometimes appears to be elsewhere. When they see students around campus, Professor S occasionally offers a wave of recognition, but does not go out of their way to engage in conversation. Professor S's responses to emails are often unclear and may appear rude in tone.
How does the faculty member you have in mind compare to Professor S?

- They are much more engaged than Professor S. (5)
- They are a little more engaged than Professor S. (4)
- They and Professor S are about the same. (3)
- Professor S is a little more engaged than them. (2)
- Professor S is much more engaged than them. (1)

Professor G is often preoccupied with other tasks during meetings with students. They do not go out of their way to greet students around campus and seem annoyed if students try to initiate conversation. However, Professor G's responses to emails are generally clear and professional.
How does the faculty member you have in mind compare to Professor G?

- They are much more engaged than Professor G. (5)
- They are a little more engaged than Professor G. (4)
- They and Professor G are about the same. (3)
- Professor G is a little more engaged than them. (2)
- Professor G is much more engaged than them. (1)

Professor R often seems annoyed during meetings with students and would clearly rather spend time on other activities. They do not acknowledge students around campus and their responses to emails are often disrespectful in tone, and do not clearly address student concerns or questions.
How does the faculty member you have in mind compare to Professor R?

- They are much more engaged than Professor R. (5)
- They are a little more engaged than Professor R. (4)

○ They and Professor R are about the same. (3)

○ Professor R is a little more engaged than them. (2)

○ Professor R is much more engaged than them. (1)

NON-OCAS ITEMS

Please estimate the number of times you have engaged in these activities with the faculty member you had in mind while completing this survey.

	1 time (1)	2 times (2)	3 times (3)	4 times (4)	5 or more times (5)
Met during scheduled office hours (1)	○	○	○	○	○
Met outside of scheduled office hours (2)	○	○	○	○	○
Informally interacted on campus (3)	○	○	○	○	○
Corresponded via email (4)	○	○	○	○	○
Had a conversation immediately before or after class (5)	○	○	○	○	○

Please check the boxes to indicate which topics you have discussed outside of class with the faculty member you had in mind while completing this survey.

☐ Grades

☐ Questions about assignment directions or expectations

- ☐ Academic advice
- ☐ Career advice
- ☐ Personal life
- ☐ Letters of recommendation
- ☐ Non-class related intellectual topics

Please indicate your overall level of satisfaction with the out-of-class communication you have had with the professor you had in mind while completing this survey.

- ○ Very satisfied (4)
- ○ Somewhat satisfied (3)
- ○ Somewhat dissatisfied (2)
- ○ Very dissatisfied (1)

What is the gender of the faculty member you had in mind while completing this survey?

____

What is the race/ethnicity of the faculty member you had in mind while completing this survey?

____

In what department is the faculty member you had in mind while completing this survey?

____

Please estimate the age of the faculty member you had in mind while completing this survey.

- ○ 30 or younger (1)
- ○ Between 30 and 39 (2)
- ○ Between 40 and 49 (3)
- ○ Between 50 and 59 (4)

○ 60 or older (5)

Appendix H: Person-Response Tables for Pre-pilot

Figure H.1

Physical Accessibility Scale Person Misfit (WINSTEPS output)

```
TABLE OF POORLY FITTING PERSON   (ITEM IN ENTRY ORDER)
NUMBER - NAME -- --------------- MEASURE - INFIT (MNSQ) OUTFIT

     10  110                        1.53     3.4   A    3.3
  OBSERVED: 1:  3    5    5    3    5     1    4    5    5
  EXPECTED:     3.1  3.9  3.1  4.0  4.4   4.2  4.4  4.3  4.6
Z-RESIDUAL:               2               -4

     32  132                        -.20     3.1   B    3.1
  OBSERVED: 1:  4    2    3    3    1     5    2    2    4
  EXPECTED:     1.9  2.8  1.9  2.8  3.3   3.1  3.3  3.2  3.7
Z-RESIDUAL:     2                  -2     2

      3  102                         .63     3.1   C    3.0
  OBSERVED: 1:  5    1    3    5    4     2    3    4    4
  EXPECTED:     2.5  3.3  2.5  3.4  3.9   3.6  3.9  3.8  4.2
Z-RESIDUAL:     2   -2                    -2

      9  109                         .30     2.3   D    2.3
  OBSERVED: 1:  4    4    3    4    2     1    4    3    4
  EXPECTED:     2.2  3.1  2.2  3.2  3.6   3.4  3.7  3.6  4.0
Z-RESIDUAL:     2                  -2    -2

     33  133                        2.78     1.6   E    1.3
  OBSERVED: 1:  5    3    3    5    5     5    5    5    5
  EXPECTED:     3.9  4.6  3.9  4.6  4.8   4.7  4.8  4.8  4.9
Z-RESIDUAL:         -2

     26  126                         .63     1.5   F    1.6
  OBSERVED: 1:  4    4    3    4    4     4    2    3    3
  EXPECTED:     2.5  3.3  2.5  3.4  3.9   3.6  3.9  3.8  4.2
Z-RESIDUAL:                                    -2

     38  138                        3.64     1.4   G     .7
  OBSERVED: 1:  3    5    5    5    5     5    5    5    5
  EXPECTED:     4.4  4.8  4.4  4.8  4.9   4.9  4.9  4.9  4.9
Z-RESIDUAL:    -2

      6  105                         .46     1.3   H    1.3
  OBSERVED: 1:  2    3    4    4    4     4    2    4    3
  EXPECTED:     2.4  3.2  2.4  3.3  3.7   3.5  3.8  3.7  4.1
Z-RESIDUAL:                                    -2

     41  141                         .30     1.3   I    1.3
  OBSERVED: 1:  3    5    3    2    3     3    4    3    3
  EXPECTED:     2.2  3.1  2.2  3.2  3.6   3.4  3.7  3.6  4.0
Z-RESIDUAL:          2

      2  101                         .13     1.3   J    1.3
  OBSERVED: 1:  4    2    3    3    3     3    4    2    4
  EXPECTED:     2.1  3.0  2.1  3.0  3.5   3.3  3.6  3.5  3.9
Z-RESIDUAL:     2

      4  103                        4.41     1.1   K    1.2
  OBSERVED: 1:  5    5    5    4    5     5    5    5    5
  EXPECTED:     4.7  4.9  4.7  4.9  5.0   4.9  5.0  5.0  5.0
Z-RESIDUAL:                   -3
```

Figure H.2

Social Engagement Scale Person Misfit (WINSTEPS output)

```
TABLE OF POORLY FITTING PERSON   (ITEM IN ENTRY ORDER)
NUMBER - NAME -- --------------- MEASURE - INFIT (MNSQ) OUTFIT

    21  121                        1.12     2.6   A    2.5
  OBSERVED: 1:  4    2    5    4    4     5    1    4    5
  EXPECTED:     2.8  3.2  3.5  4.0  4.4   4.0  3.8  4.1  4.2
Z-RESIDUAL:                                    -3

    32  132                         .95     2.2   B    2.6
  OBSERVED: 1:  4    4    4    2    2     4    4    4    5
  EXPECTED:     2.6  3.0  3.4  3.9  4.3   3.9  3.7  4.0  4.1
Z-RESIDUAL:                   -2   -3

     9  109                        1.50     2.2   C    2.1
  OBSERVED: 1:  5    5    4    3    5     4    4    3    3
  EXPECTED:     3.1  3.5  3.8  4.2  4.5   4.2  4.0  4.3  4.4
Z-RESIDUAL:     2                                        -2

     3  102                         .15     1.8   D    1.7
  OBSERVED: 1:  3    3    3    4    3     5    1    2    4
  EXPECTED:     2.1  2.4  2.8  3.3  3.9   3.4  3.1  3.5  3.6
Z-RESIDUAL:                                    -2

    10  110                         .78     1.5   E    1.6
  OBSERVED: 1:  1    4    4    4    5     4    4    4    2
  EXPECTED:     2.5  2.9  3.3  3.7  4.2   3.8  3.6  3.9  4.0
Z-RESIDUAL:                                              -2

    12  112                         .30     1.4   F    1.4
  OBSERVED: 1:  1    4    4    4    4     2    2    4    4
  EXPECTED:     2.2  2.5  2.9  3.4  4.0   3.5  3.2  3.6  3.7
Z-RESIDUAL:

    37  137                        -.15     1.3   G    1.4
  OBSERVED: 1:  1    1    2    3    5     5    2    3    4
  EXPECTED:     1.9  2.2  2.5  3.0  3.7   3.1  2.8  3.2  3.4
Z-RESIDUAL:                               2

    41  141                         .15     1.3   H    1.3
  OBSERVED: 1:  3    2    4    3    4     2    3    2    5
  EXPECTED:     2.1  2.4  2.8  3.3  3.9   3.4  3.1  3.5  3.6
Z-RESIDUAL:

     6  105                         .15     1.3   I    1.2
  OBSERVED: 1:  2    2    4    4    4     4    4    2    2
  EXPECTED:     2.1  2.4  2.8  3.3  3.9   3.4  3.1  3.5  3.6
Z-RESIDUAL:
```

Appendix I: Physical Accessibility Rasch Results when Misfitting Individual 122 is Included

Figure I.1

Physical Accessibility Scale Variable Map Including Individual 122 (Pilot)

```
MEASURE                        PERSON - MAP - ITEM
                                   <more>|<rare>
    6                               164  +
                                         |
                                         |
                                         |
                                         |
    5                                    +
                                        T|
                                         |
                                         |
                               154  158  |
    4                                    +
           132  134  149  152  159  160  |
                                         |
      127  133  136  137  147  150  156 S|T
                                         |
    3      109  110  135  138  151  155  +  HHH
                                         |
                115  142  143  161  162  |
                     111  123  144  157  |  HMH
                                         |
    2           130  131  146  148  153 M+
           103  114  120  124  125  128  |S
                     104  105  119  163  |
                                         |
                               121  140  |
    1                          108  129  +
                               117  126  |
                               141  145 S|  HHM    MMH
                               116  139  |
                                    101  |
    0                          112  118  +M
                                    107  |
                               102  113  |
                                    106  |  MLM
                                        T|  MMM
   -1                                    +  MML
                                         |
                                         |
                                         |
                                         |S
   -2                               122  +  LLL
                                         |  LLM
                                         |
                                         |
                                         |
   -3                                    +
                                   <less>|<freq>
```

Table I.1

Physical Accessibility Fit Statistics Including Individual 122 (Pilot)

Item	Logit Estimate (S.E.)	Information-Weighted Fit Statistic		Unweighted Fit Statistic	
		MNSQ	ZSTD	MNSQ	ZSTD
Item 1: M-H, I-H, E-H	3.03 (.17)	.91	-.48	1.13	.77
Item 3: M-H, I-M, E-H	2.44 (.17)	1.00	.05	1.86	4.04
Item 2: M-H, I-H, E-M	.70 (.18)	1.20	1.15	1.29	1.59
Item 4: M-M, I-M, E-H	.57 (.18)	.71	-1.77	.72	-1.71
Item 6: M-M, I-L, E-M	-.55 (.21)	.96	-.17	.81	-.91
Item 5: M-M, I-M, E-M	-.87 (.22)	.99	.01	.87	-.52
Item 7: M-M, I-M, E-L	-1.01 (.22)	.97	-.10	.70	-1.30
Item 9: M-L, I-L, E-L	-2.07 (.28)	1.57	2.00	1.52	1.29
Item 8: M-L, I-L, E-M	-2.23 (.30)	1.29	1.10	.91	-.07

Figure I.2

Individual 122's Response Pattern

```
22  122                       -2.09     7.8   A    9.9
  OBSERVED: 1:  2    4    4    3    1     1    1    1    1
  EXPECTED:     1.1  1.5  1.1  1.6  2.2   2.1  2.3  3.1  3.0
Z-RESIDUAL:     3    4    8    2                   -2   -2
```

Appendix J: Person-Response Tables from Pilot

Figure J.1

Physical Accessibility Scale Person Misfit (WINSTEPS output)

```
TABLE OF POORLY FITTING PERSON   (ITEM IN ENTRY ORDER)
NUMBER - NAME -- --------------- MEASURE - INFIT (MNSQ) OUTFIT

    11  111                        2.69     3.8   A    9.9
  OBSERVED: 1:  2    5    3    5    5    5    5    5    2
  EXPECTED:     2.5  3.9  2.9  4.0  4.6   4.5  4.7  4.9  4.9
Z-RESIDUAL:                                            -9

    18  118                        -.04     5.3   B    4.8
  OBSERVED: 1:  2    4    2    4    5    1    1    3    5
  EXPECTED:     1.4  2.5  1.6  2.6  3.5   3.3  3.6  4.3  4.2
Z-RESIDUAL:          2              2    -3   -3   -2

    62  163                        1.77     3.7   C    3.7
  OBSERVED: 1:  4    4    4    4    4    2    4    4    4
  EXPECTED:     2.1  3.5  2.4  3.5  4.3   4.2  4.4  4.8  4.8
Z-RESIDUAL:     2         2              -3

    59  160                        4.31     2.3   D    1.4
  OBSERVED: 1:  3    3    5    5    5    5    5    5    5
  EXPECTED:     3.4  4.6  3.7  4.6  4.9   4.9  4.9  5.0  5.0
Z-RESIDUAL:         -2

    41  142                        3.04      .9   E    2.3
  OBSERVED: 1:  3    4    4    4    4    5    5    4    5
  EXPECTED:     2.7  4.1  3.0  4.2  4.7   4.6  4.8  4.9  4.9
Z-RESIDUAL:                                       -3

    61  162                        3.04     2.3   F    1.6
  OBSERVED: 1:  4    2    3    4    5    5    5    5    5
  EXPECTED:     2.7  4.1  3.0  4.2  4.7   4.6  4.8  4.9  4.9
Z-RESIDUAL:         -3

    42  143                        3.04      .9   G    2.1
  OBSERVED: 1:  3    4    4    4    5    4    5    4    5
  EXPECTED:     2.7  4.1  3.0  4.2  4.7   4.6  4.8  4.9  4.9
Z-RESIDUAL:                                       -3

    40  141                         .70     2.0   H    1.9
  OBSERVED: 1:  3    1    2    2    5    4    4    5    4
  EXPECTED:     1.7  2.9  1.9  3.0  3.9   3.7  4.0  4.6  4.5
Z-RESIDUAL:     2   -2

     3  103                        2.06     1.5   I    1.3
  OBSERVED: 1:  1    4    1    4    5    5    5    5    5
  EXPECTED:     2.2  3.6  2.5  3.7  4.5   4.3  4.5  4.8  4.8
Z-RESIDUAL:              -2

    43  144                        2.69      .6   J    1.5
  OBSERVED: 1:  3    4    3    4    4    5    5    4    5
  EXPECTED:     2.5  3.9  2.9  4.0  4.6   4.5  4.7  4.9  4.9
Z-RESIDUAL:                                       -3

    52  153                        2.36     1.2   K    1.4
  OBSERVED: 1:  3    5    3    4    4    4    4    5    4
  EXPECTED:     2.4  3.8  2.7  3.8  4.6   4.4  4.6  4.9  4.9
Z-RESIDUAL:                                            -2

    30  131                        2.36     1.4   L    1.1
```

```
  OBSERVED: 1:  3    3    4    3    4     5    4    5    5
  EXPECTED:     2.4  3.8  2.7  3.8  4.6   4.4  4.6  4.9  4.9
Z-RESIDUAL:

     38  139                         .45     1.3   M    1.4
  OBSERVED: 1:  3    2    2    4    3     3    4    4    4
  EXPECTED:     1.6  2.7  1.8  2.8  3.8   3.6  3.8  4.5  4.4
Z-RESIDUAL:     2

     60  161                        3.04     1.3   N    1.1
  OBSERVED: 1:  4    4    3    3    4     5    5    5    5
  EXPECTED:     2.7  4.1  3.0  4.2  4.7   4.6  4.8  4.9  4.9
Z-RESIDUAL:

     36  137                        3.83     1.3   O     .7
  OBSERVED: 1:  3    5    2    5    5     5    5    5    5
  EXPECTED:     3.2  4.4  3.5  4.5  4.9   4.8  4.9  5.0  5.0
Z-RESIDUAL:              -2

     13  113                        -.53     1.3   P    1.3
  OBSERVED: 1:  2    3    1    3    4     2    2    4    4
  EXPECTED:     1.3  2.2  1.5  2.3  3.2   3.0  3.3  4.1  4.0
Z-RESIDUAL:

     34  135                        3.41     1.0   Q    1.3
  OBSERVED: 1:  3    5    3    5    4     5    4    5    5
  EXPECTED:     2.9  4.2  3.3  4.3  4.8   4.7  4.8  5.0  4.9
Z-RESIDUAL:                                   -2

      1  101                         .21     1.2   R    1.2
  OBSERVED: 1:  1    3    2    3    5     3    4    3    4
  EXPECTED:     1.5  2.6  1.7  2.7  3.6   3.4  3.7  4.4  4.3
Z-RESIDUAL:                                        -2
```

Figure J.2

Social Engagement Scale Person Misfit (WINSTEPS output)

```
TABLE OF POORLY FITTING PERSON   (ITEM IN ENTRY ORDER)
NUMBER - NAME -- --------------- MEASURE - INFIT (MNSQ) OUTFIT

     29  133                        4.59     1.2   A    9.9
  OBSERVED: 1:  3    5    5    5    5     5    5    5    4
  EXPECTED:     3.6  4.4  4.6  4.7  4.9   4.9  4.9  4.9  5.0
Z-RESIDUAL:                                             -9

     16  117                         .27     6.0   B    5.9
  OBSERVED: 1:  5    3    2    2    4     M    4    1    4
  EXPECTED:     1.6  2.5  2.7  3.0  3.4        3.6  3.7  4.5
Z-RESIDUAL:     5                                  -3

     30  135                        2.01     4.3   C    3.9
  OBSERVED: 1:  5    3    5    3    3     5    3    3    5
  EXPECTED:     2.5  3.2  3.5  3.7  4.2   4.1  4.4  4.5  4.9
Z-RESIDUAL:     3         2                   -2   -2

     34  140                         .21     2.7   D    2.6
  OBSERVED: 1:  2    3    2    3    3     4    1    5    5
  EXPECTED:     1.6  2.4  2.7  3.0  3.4   3.2  3.6  3.6  4.5
Z-RESIDUAL:                                   -3

      2  102                         .46     2.5   E    2.4
  OBSERVED: 1:  1    3    3    4    4     4    1    4    5
  EXPECTED:     1.7  2.6  2.8  3.1  3.5   3.3  3.7  3.8  4.5
```

```
Z-RESIDUAL:                                   -3

     22  125                        1.48     2.3   F    2.1
  OBSERVED: 1:  1    4    4    2    4     5    3    5    5
  EXPECTED:     2.2  3.0  3.3  3.5  4.0   3.8  4.2  4.2  4.8
Z-RESIDUAL:                   -2


     32  138                         .46     2.2   G    2.1
  OBSERVED: 1:  4    3    3    3    2     3    4    3    4
  EXPECTED:     1.7  2.6  2.8  3.1  3.5   3.3  3.7  3.8  4.5
Z-RESIDUAL:     3                  -2

     52  162                        -.30     1.9   H    2.1
  OBSERVED: 1:  3    3    3    3    3     3    3    2    3
  EXPECTED:     1.4  2.2  2.5  2.7  3.2   3.0  3.4  3.4  4.3
Z-RESIDUAL:     2                                  -2

     11  111                        -.30     1.5   I    1.6
  OBSERVED: 1:  2    2    3    1    3     2    4    4    5
  EXPECTED:     1.4  2.2  2.5  2.7  3.2   3.0  3.4  3.4  4.3
Z-RESIDUAL:                   -2

     10  110                        3.63     1.5   J    1.5
  OBSERVED: 1:  2    5    4    5    5     5    5    4    5
  EXPECTED:     3.2  4.0  4.2  4.5  4.8   4.7  4.8  4.9  5.0
Z-RESIDUAL:                                        -2

     27  130                        1.74     1.5   K    1.4
  OBSERVED: 1:  2    4    4    2    4     3    5    5    5
  EXPECTED:     2.4  3.1  3.4  3.6  4.1   3.9  4.3  4.4  4.8
Z-RESIDUAL:                   -2

     42  151                        4.07     1.5   L     .9
  OBSERVED: 1:  3    5    3    5    5     5    5    5    5
  EXPECTED:     3.4  4.2  4.4  4.6  4.8   4.8  4.9  4.9  5.0
Z-RESIDUAL:              -2
```

Appendix K: Person-Response Tables from Final Administration

Figure K.1

Physical Accessibility Scale Person Misfit (WINSTEPS output)

```
TABLE OF POORLY FITTING PERSON   (ITEM IN ENTRY ORDER)
NUMBER - NAME -- --------------- MEASURE - INFIT (MNSQ) OUTFIT

     95   95                       -1.77     6.8   A    9.9
  OBSERVED: 1:  4    1    4    2    1     1    1
  EXPECTED:     1.1  1.7  1.5  2.0  2.2   2.7  2.8
Z-RESIDUAL:     8         4               -2   -2

    152  152                       -2.42     4.2   B    9.9
  OBSERVED: 1:  3    2    3    1    1     1    1
  EXPECTED:     1.1  1.4  1.3  1.7  1.9   2.3  2.4
Z-RESIDUAL:     7         3

     50   50                       -1.77     6.1   C    9.6
  OBSERVED: 1:  3    3    4    1    1     1    1
  EXPECTED:     1.1  1.7  1.5  2.0  2.2   2.7  2.8
Z-RESIDUAL:     5         4               -2   -2

    157  157                       -1.77     5.7   D    9.2
  OBSERVED: 1:  3    1    4    3    1     1    1
  EXPECTED:     1.1  1.7  1.5  2.0  2.2   2.7  2.8
Z-RESIDUAL:     5         4               -2   -2

     97   97                       -1.48     6.1   E    8.2
  OBSERVED: 1:  3    5    1    2    2     1    1
  EXPECTED:     1.2  1.8  1.6  2.1  2.4   2.9  3.0
Z-RESIDUAL:     4    4                    -2   -2

    156  156                         .71     6.5   F    7.0
  OBSERVED: 1:  5    2    5    2    5     2    3
  EXPECTED:     1.9  3.1  2.8  3.6  3.9   4.4  4.4
Z-RESIDUAL:     4         2               -3

      1    1                        -.70     5.0   G    6.9
  OBSERVED: 1:  4    3    1    1    5     2    2
  EXPECTED:     1.3  2.2  2.0  2.6  2.9   3.5  3.5
Z-RESIDUAL:     5              -2    2

     93   93                        -.22     4.7   H    6.8
  OBSERVED: 1:  5    2    3    3    3     2    2
  EXPECTED:     1.4  2.5  2.3  2.9  3.2   3.8  3.8
Z-RESIDUAL:     6                         -2   -2

    190  190                       -2.42     2.6   I    4.7
  OBSERVED: 1:  2    2    3    1    1     2    1
  EXPECTED:     1.1  1.4  1.3  1.7  1.9   2.3  2.4
Z-RESIDUAL:     3         3

     41   41                        1.20     4.6   J    4.2
  OBSERVED: 1:  3    2    5    5    1     5    5
  EXPECTED:     2.1  3.5  3.2  3.9  4.2   4.6  4.6
Z-RESIDUAL:               2         -4

    206  207                         .01     3.5   K    3.7
  OBSERVED: 1:  3    2    4    3    3     1    5
  EXPECTED:     1.5  2.7  2.4  3.1  3.4   4.0  4.0
Z-RESIDUAL:     2         2               -3

    137  137                        2.05     3.7   L    2.7
  OBSERVED: 1:  3    1    5    5    5     5    5
```

```
  EXPECTED:     2.6  4.0  3.8  4.4  4.6   4.8  4.8
Z-RESIDUAL:         -3

    192  192                       -1.48     3.0   M    3.6
  OBSERVED: 1:  2    2    4    1    1     2    3
  EXPECTED:     1.2  1.8  1.6  2.1  2.4   2.9  3.0
Z-RESIDUAL:     2         3

     40   40                         .24     3.5   N    3.4
  OBSERVED: 1:  1    1    5    5    2     4    4
  EXPECTED:     1.6  2.8  2.5  3.2  3.5   4.1  4.1
Z-RESIDUAL:         -2    3    2

     13   13                         .01     2.2   O    2.9
  OBSERVED: 1:  4    3    2    3    2     4    3
  EXPECTED:     1.5  2.7  2.4  3.1  3.4   4.0  4.0
Z-RESIDUAL:     3

    159  159                        -.95     2.8   P    2.7
  OBSERVED: 1:  1    5    2    2    2     2    3
  EXPECTED:     1.2  2.1  1.9  2.5  2.7   3.3  3.3
Z-RESIDUAL:          3

    228  230                         .24     2.6   Q    2.8
  OBSERVED: 1:  3    2    4    4    4     3    2
  EXPECTED:     1.6  2.8  2.5  3.2  3.5   4.1  4.1
Z-RESIDUAL:     2                              -2

     83   83                        -.95     2.5   R    2.4
  OBSERVED: 1:  2    3    1    2    4     1    4
  EXPECTED:     1.2  2.1  1.9  2.5  2.7   3.3  3.3
Z-RESIDUAL:                              -2

     16   16                        1.75     2.3   S    1.9
  OBSERVED: 1:  4    2    4    3    5     5    5
  EXPECTED:     2.4  3.8  3.5  4.2  4.5   4.7  4.7
Z-RESIDUAL:         -2

     25   25                        2.38     2.3   T    1.7
  OBSERVED: 1:  3    2    5    5    5     5    5
  EXPECTED:     2.8  4.2  4.0  4.5  4.7   4.9  4.9
Z-RESIDUAL:         -2

    178  178                        3.82     1.8   U    1.5
  OBSERVED: 1:  5    5    4    4    5     5    5
  EXPECTED:     3.8  4.8  4.7  4.9  4.9   5.0  5.0
Z-RESIDUAL:                   -2

    200  200                        1.47     1.8   V    1.6
  OBSERVED: 1:  3    2    5    4    4     5    4
  EXPECTED:     2.3  3.6  3.4  4.1  4.3   4.7  4.7
Z-RESIDUAL:

     38   38                         .71     1.7   W    1.8
  OBSERVED: 1:  3    2    4    4    3     3    5
  EXPECTED:     1.9  3.1  2.8  3.6  3.9   4.4  4.4
Z-RESIDUAL:

    146  146                        2.38     1.7   X    1.2
  OBSERVED: 1:  3    5    2    5    5     5    5
  EXPECTED:     2.8  4.2  4.0  4.5  4.7   4.9  4.9
Z-RESIDUAL:              -2

    181  181                        1.47     1.7   Y    1.5
  OBSERVED: 1:  1    2    4    5    5     5    5
  EXPECTED:     2.3  3.6  3.4  4.1  4.3   4.7  4.7
Z-RESIDUAL:

     61   61                         .24     1.7   Z    1.7
```

```
  OBSERVED: 1:  2    4    3    4    3     4    2
  EXPECTED:     1.6  2.8  2.5  3.2  3.5   4.1  4.1
Z-RESIDUAL:                                   -2

    132  132                         .71     1.6        1.6
  OBSERVED: 1:  2    5    3    3    3     3    5
  EXPECTED:     1.9  3.1  2.8  3.6  3.9   4.4  4.4
Z-RESIDUAL:          2

    229  231                         .61     1.6        1.5
  OBSERVED: 1:  1    4    M    2    M     5    5
  EXPECTED:     1.8  3.1       3.5        4.3  4.3
Z-RESIDUAL:

    150  150                         .48     1.5        1.5
  OBSERVED: 1:  2    5    2    3    4     4    3
  EXPECTED:     1.7  3.0  2.7  3.4  3.7   4.2  4.3
Z-RESIDUAL:          2

    103  103                         .48     1.5        1.5
  OBSERVED: 1:  2    5    3    3    3     3    4
  EXPECTED:     1.7  3.0  2.7  3.4  3.7   4.2  4.3
Z-RESIDUAL:          2

     92   92                        2.05     1.5        1.2
  OBSERVED: 1:  1    5    3    5    5     5    5
  EXPECTED:     2.6  4.0  3.8  4.4  4.6   4.8  4.8
Z-RESIDUAL:    -2

    118  118                        2.38     1.5        1.1
  OBSERVED: 1:  4    3    3    5    5     5    5
  EXPECTED:     2.8  4.2  4.0  4.5  4.7   4.9  4.9
Z-RESIDUAL:

     10   10                        4.71     1.5        1.4
  OBSERVED: 1:  5    4    5    5    5     5    5
  EXPECTED:     4.4  4.9  4.9  4.9  5.0   5.0  5.0
Z-RESIDUAL:         -2

    110  110                         .95     1.5        1.4
  OBSERVED: 1:  1    4    3    2    5     5    5
  EXPECTED:     2.0  3.3  3.0  3.7  4.0   4.5  4.5
Z-RESIDUAL:                   -2

    149  149                         .48     1.4        1.4
  OBSERVED: 1:  1    1    3    4    4     5    5
  EXPECTED:     1.7  3.0  2.7  3.4  3.7   4.2  4.3
Z-RESIDUAL:         -2

    134  134                        3.23     1.4         .9
  OBSERVED: 1:  4    5    3    5    5     5    5
  EXPECTED:     3.4  4.6  4.4  4.8  4.9   4.9  4.9
Z-RESIDUAL:              -2

    187  187                        4.71     1.4         .9
  OBSERVED: 1:  5    5    4    5    5     5    5
  EXPECTED:     4.4  4.9  4.9  4.9  5.0   5.0  5.0
Z-RESIDUAL:              -2

    231  233                        1.20     1.4        1.2
  OBSERVED: 1:  3    2    2    4    5     5    5
  EXPECTED:     2.1  3.5  3.2  3.9  4.2   4.6  4.6
Z-RESIDUAL:

     30   30                         .71     1.3        1.3
  OBSERVED: 1:  1    2    3    5    3     5    5
  EXPECTED:     1.9  3.1  2.8  3.6  3.9   4.4  4.4
Z-RESIDUAL:
```

```
      2    2                     1.04     1.1        1.3
  OBSERVED: 1:  3    M    3    4    4     3    5
  EXPECTED:     2.0       3.1  3.8  4.1   4.5  4.5
Z-RESIDUAL:                               -2

    101  101                     1.20     1.3        1.2
  OBSERVED: 1:  1    2    4    4    5     5    5
  EXPECTED:     2.1  3.5  3.2  3.9  4.2   4.6  4.6
Z-RESIDUAL:

     62   62                     2.38     1.3        1.0
  OBSERVED: 1:  2    3    5    5    5     5    5
  EXPECTED:     2.8  4.2  4.0  4.5  4.7   4.9  4.9
Z-RESIDUAL:

      3    3                      .95     1.3        1.2
  OBSERVED: 1:  2    4    3    4    2     5    5
  EXPECTED:     2.0  3.3  3.0  3.7  4.0   4.5  4.5
Z-RESIDUAL:                         -2

     33   33                     1.75     1.3        1.3
  OBSERVED: 1:  3    5    3    3    5     4    5
  EXPECTED:     2.4  3.8  3.5  4.2  4.5   4.7  4.7
Z-RESIDUAL:

    176  176                     1.20     1.3        1.2
  OBSERVED: 1:  2    4    3    2    5     5    5
  EXPECTED:     2.1  3.5  3.2  3.9  4.2   4.6  4.6
Z-RESIDUAL:                    -2

     39   39                     1.20     1.2        1.1
  OBSERVED: 1:  3    2    4    3    4     5    5
  EXPECTED:     2.1  3.5  3.2  3.9  4.2   4.6  4.6
Z-RESIDUAL:

    203  204                     1.20     1.2        1.1
  OBSERVED: 1:  1    5    3    3    4     5    5
  EXPECTED:     2.1  3.5  3.2  3.9  4.2   4.6  4.6
Z-RESIDUAL:

     89   89                     1.47     1.2        1.0
  OBSERVED: 1:  2    2    3    5    5     5    5
  EXPECTED:     2.3  3.6  3.4  4.1  4.3   4.7  4.7
Z-RESIDUAL:

    148  148                     1.47     1.2        1.0
  OBSERVED: 1:  2    2    3    5    5     5    5
  EXPECTED:     2.3  3.6  3.4  4.1  4.3   4.7  4.7
Z-RESIDUAL:

     15   15                     2.05     1.2        1.0
  OBSERVED: 1:  2    3    5    5    4     5    5
  EXPECTED:     2.6  4.0  3.8  4.4  4.6   4.8  4.8
Z-RESIDUAL:
```

Figure K.2

Social Engagement Scale Person Misfit (WINSTEPS output)

```
TABLE OF POORLY FITTING PERSON   (ITEM IN ENTRY ORDER)
NUMBER - NAME -- --------------- MEASURE - INFIT (MNSQ) OUTFIT

      3  152                    -3.91     4.3   A    9.9
  OBSERVED: 1:  3    2    2    1    1     1    1
  EXPECTED:     1.0  1.1  1.3  1.6  1.8   1.9  2.4
Z-RESIDUAL:     9    4
```

```
      2  158                      -5.04     1.6   B    8.5
  OBSERVED: 1:  1    2    1    1    1     2    1
  EXPECTED:     1.0  1.0  1.1  1.2  1.4   1.4  1.8
Z-RESIDUAL:          7

    107   93                       2.00     8.3   C    8.1
  OBSERVED: 1:  5    4    5    4    3     1    5
  EXPECTED:     2.5  2.9  3.8  4.2  4.4   4.4  4.8
Z-RESIDUAL:     3                   -2    -5

      6  207                       1.00     6.4   D    6.6
  OBSERVED: 1:  5    2    5    4    3     3    2
  EXPECTED:     2.0  2.5  3.3  3.7  4.0   4.0  4.5
Z-RESIDUAL:     4         2                    -4

     43   16                       2.38     5.7   E    5.2
  OBSERVED: 1:  5    4    4    2    3     5    5
  EXPECTED:     2.7  3.1  4.0  4.3  4.6   4.6  4.8
Z-RESIDUAL:     3              -3   -2

      4  132                        .04     5.4   F    5.2
  OBSERVED: 1:  3    2    4    2    5     4    1
  EXPECTED:     1.6  2.1  3.0  3.3  3.5   3.5  4.0
Z-RESIDUAL:     2                   2          -4

     41   83                        .36     4.2   G    4.3
  OBSERVED: 1:  1    4    5    2    2     3    5
  EXPECTED:     1.7  2.2  3.1  3.4  3.7   3.7  4.2
Z-RESIDUAL:          2    3   -2   -2

    183  155                       5.28     1.2   H    4.3
  OBSERVED: 1:  4    5    5    5    4     5    5
  EXPECTED:     3.9  4.4  4.9  4.9  5.0   5.0  5.0
Z-RESIDUAL:                         -5

    184  194                       5.28     1.2   I    4.3
  OBSERVED: 1:  4    5    5    5    4     5    5
  EXPECTED:     3.9  4.4  4.9  4.9  5.0   5.0  5.0
Z-RESIDUAL:                         -5

      9  103                       1.65     4.0   J    4.3
  OBSERVED: 1:  4    4    4    2    4     5    3
  EXPECTED:     2.3  2.8  3.6  4.0  4.3   4.3  4.7
Z-RESIDUAL:     2              -2              -3

    185  140                       3.84     3.9   K    2.2
  OBSERVED: 1:  5    2    4    5    5     5    5
  EXPECTED:     3.2  3.7  4.6  4.8  4.9   4.9  5.0
Z-RESIDUAL:     2   -2

      5  150                        .68     3.1   L    3.2
  OBSERVED: 1:  3    2    4    4    5     3    2
  EXPECTED:     1.9  2.4  3.2  3.5  3.8   3.9  4.3
Z-RESIDUAL:                                    -3

     14   13                        .04     3.2   M    3.2
  OBSERVED: 1:  3    4    3    2    2     3    4
  EXPECTED:     1.6  2.1  3.0  3.3  3.5   3.5  4.0
Z-RESIDUAL:     2    2              -2

     39  178                       2.38     2.9   N    3.2
  OBSERVED: 1:  4    4    3    5    5     3    4
  EXPECTED:     2.7  3.1  4.0  4.3  4.6   4.6  4.8
Z-RESIDUAL:     2                         -2   -2

    109   92                       1.00     2.9   O    2.9
  OBSERVED: 1:  1    1    2    5    5     5    5
  EXPECTED:     2.0  2.5  3.3  3.7  4.0   4.0  4.5
```

```
Z-RESIDUAL:         -2   -2

     45  198                        1.00     2.8   P    2.9
  OBSERVED: 1:  1    5    3    3    3     4    5
  EXPECTED:     2.0  2.5  3.3  3.7  4.0   4.0  4.5
Z-RESIDUAL:          3

    113   29                        3.28     2.8   Q    2.0
  OBSERVED: 1:  2    5    3    5    5     5    5
  EXPECTED:     3.0  3.5  4.4  4.6  4.8   4.8  4.9
Z-RESIDUAL:          2   -2

     25   75                        2.80     2.2   R    2.5
  OBSERVED: 1:  2    5    4    4    5     5    4
  EXPECTED:     2.8  3.3  4.2  4.5  4.7   4.7  4.9
Z-RESIDUAL:          2                        -2

      8    2                         .36     2.2   S    2.3
  OBSERVED: 1:  3    3    4    2    4     3    3
  EXPECTED:     1.7  2.2  3.1  3.4  3.7   3.7  4.2
Z-RESIDUAL:                   -2

     98  143                        3.28     2.3   T    2.1
  OBSERVED: 1:  3    5    5    4    4     4    5
  EXPECTED:     3.0  3.5  4.4  4.6  4.8   4.8  4.9
Z-RESIDUAL:          2

    170   45                        4.49     1.5   U    2.3
  OBSERVED: 1:  3    5    5    5    4     5    5
  EXPECTED:     3.5  4.0  4.8  4.9  4.9   4.9  5.0
Z-RESIDUAL:                        -3

      7  192                        -.65     2.1   V    2.1
  OBSERVED: 1:  1    2    2    2    5     4    3
  EXPECTED:     1.4  1.8  2.7  3.0  3.2   3.3  3.7
Z-RESIDUAL:                         2

     11  101                        1.00     1.9   W    2.0
  OBSERVED: 1:  1    2    4    4    5     5    3
  EXPECTED:     2.0  2.5  3.3  3.7  4.0   4.0  4.5
Z-RESIDUAL:                                   -2

     35   38                        1.65     2.0   X    2.0
  OBSERVED: 1:  4    3    3    4    5     3    4
  EXPECTED:     2.3  2.8  3.6  4.0  4.3   4.3  4.7
Z-RESIDUAL:     2

     40   33                        2.07     1.8   Y    1.7
  OBSERVED: 1:  3    3    5    M    4     3    5
  EXPECTED:     2.5  3.0  3.8       4.4   4.5  4.8
Z-RESIDUAL:                              -2

     13   40                         .36     1.7   Z    1.7
  OBSERVED: 1:  1    2    4    2    4     5    4
  EXPECTED:     1.7  2.2  3.1  3.4  3.7   3.7  4.2
Z-RESIDUAL:                   -2

     38  164                        2.38     1.4        1.6
  OBSERVED: 1:  3    4    3    5    5     4    4
  EXPECTED:     2.7  3.1  4.0  4.3  4.6   4.6  4.8
Z-RESIDUAL:                                   -2

    110  234                        2.00     1.5        1.4
  OBSERVED: 1:  1    3    3    5    5     5    5
  EXPECTED:     2.5  2.9  3.8  4.2  4.4   4.4  4.8
Z-RESIDUAL:    -2

     12   67                        -.30     1.4        1.5
  OBSERVED: 1:  1    2    3    2    3     5    4
```

```
  EXPECTED:      1.5  1.9  2.8  3.1  3.4   3.4  3.9
Z-RESIDUAL:                                2

    108   71                        1.65     1.5        1.4
  OBSERVED: 1:  1    2    4    5    5     4    5
  EXPECTED:     2.3  2.8  3.6  4.0  4.3   4.3  4.7
Z-RESIDUAL:

     50   44                        2.80     1.4        1.3
  OBSERVED: 1:  3    3    5    3    5     5    5
  EXPECTED:     2.8  3.3  4.2  4.5  4.7   4.7  4.9
Z-RESIDUAL:                   -2

    168   41                        3.28     1.4        1.0
  OBSERVED: 1:  3    2    5    5    5     5    5
  EXPECTED:     3.0  3.5  4.4  4.6  4.8   4.8  4.9
Z-RESIDUAL:         -2

     49  221                        2.38     1.4        1.2
  OBSERVED: 1:  3    3    5    3    4     5    5
  EXPECTED:     2.7  3.1  4.0  4.3  4.6   4.6  4.8
Z-RESIDUAL:                   -2

     28   91                        2.38     1.0        1.3
  OBSERVED: 1:  2    4    4    4    5     5    4
  EXPECTED:     2.7  3.1  4.0  4.3  4.6   4.6  4.8
Z-RESIDUAL:                                   -2

     36  142                        2.00     1.2        1.3
  OBSERVED: 1:  4    3    4    4    4     4    4
  EXPECTED:     2.5  2.9  3.8  4.2  4.4   4.4  4.8
Z-RESIDUAL:     2

    145   31                        2.38     1.2        1.1
  OBSERVED: 1:  3    2    3    5    5     5    5
  EXPECTED:     2.7  3.1  4.0  4.3  4.6   4.6  4.8
Z-RESIDUAL:
```

Appendix L: Score Conversion Tables (Final Data)

Figure L.1

Physical Accessibility Scale Score Conversion Table (WINSTEPS output)

TABLE OF SAMPLE NORMS (500/100) AND FREQUENCIES CORRESPONDING TO COMPLETE TEST

SCORE	MEASURE	S.E.	NORMED	S.E.	FREQUENCY	%	CUM.FREQ.	%	PERCENTILE
7	-5.97E	1.87	-1	118	1	.4	1	.4	1
8	-4.66	1.07	82	68	0	.0	1	.4	1
9	-3.81	.81	135	51	0	.0	1	.4	1
10	-3.25	.70	171	44	0	.0	1	.4	1
11	-2.80	.64	199	40	0	.0	1	.4	1
12	-2.42	.60	223	38	2	.9	3	1.3	1
13	-2.09	.57	245	36	0	.0	3	1.3	1
14	-1.77	.55	265	35	3	1.3	6	2.6	2
15	-1.49	.53	283	34	4	1.7	10	4.3	3
16	-1.21	.52	300	33	1	.4	11	4.7	4
17	-.95	.50	317	32	3	1.3	14	6.0	5
18	-.70	.50	333	31	2	.9	16	6.8	6
19	-.46	.49	348	31	0	.0	16	6.8	7
20	-.22	.48	363	31	3	1.3	19	8.1	7
21	.01	.48	378	31	4	1.7	23	9.8	9
22	.24	.48	393	31	6	2.6	29	12.3	11
23	.48	.48	407	31	4	1.7	33	14.0	13
24	.71	.49	422	31	10	4.3	43	18.3	16
25	.96	.50	438	31	11	4.7	54	23.0	21
26	1.21	.51	454	32	14	6.0	68	28.9	26
27	1.47	.52	470	33	18	7.7	86	36.6	33
28	1.75	.54	488	34	23	9.8	109	46.4	41
29	2.05	.56	507	36	18	7.7	127	54.0	50
30	2.39	.60	529	38	27	11.5	154	65.5	60
31	2.77	.65	553	41	34	14.5	188	80.0	73
32	3.23	.72	582	46	23	9.8	211	89.8	85
33	3.83	.83	620	53	15	6.4	226	96.2	93
34	4.71	1.09	676	69	2	.9	228	97.0	97
35	6.04E	1.87	760	119	7	3.0	235	100.0	99

Figure L.2

Social Engagement Scale Score Conversion Table (WINSTEPS output)

TABLE OF SAMPLE NORMS (500/100) AND FREQUENCIES CORRESPONDING TO COMPLETE TEST

SCORE	MEASURE	S.E.	NORMED	S.E.	FREQUENCY	%	CUM.FREQ.	%	PERCENTILE
7	-7.31E	1.88	-57	107	0	.0	0	.0	0
8	-5.96	1.10	20	63	0	.0	0	.0	0
9	-5.05	.85	72	48	1	.5	1	.5	1
10	-4.42	.74	107	42	0	.0	1	.5	1
11	-3.91	.69	136	39	1	.5	2	1.0	1
12	-3.45	.66	162	38	0	.0	2	1.0	1
13	-3.02	.65	187	37	0	.0	2	1.0	1
14	-2.60	.64	211	37	0	.0	2	1.0	1
15	-2.19	.64	234	36	1	.5	3	1.6	1
16	-1.79	.63	257	36	0	.0	3	1.6	2
17	-1.40	.62	279	35	0	.0	3	1.6	2
18	-1.01	.61	301	35	1	.5	4	2.1	2
19	-.65	.60	322	34	1	.5	5	2.6	2
20	-.30	.58	342	33	2	1.0	7	3.7	3
21	.04	.57	361	33	5	2.6	12	6.3	5
22	.36	.57	379	32	7	3.7	19	9.9	8
23	.68	.56	397	32	8	4.2	27	14.1	12
24	1.00	.56	415	32	13	6.8	40	20.9	18
25	1.32	.57	434	32	4	2.1	44	23.0	22
26	1.65	.58	453	33	14	7.3	58	30.4	27
27	2.00	.60	473	34	15	7.9	73	38.2	34
28	2.38	.63	494	36	30	15.7	103	53.9	46
29	2.80	.67	518	38	22	11.5	125	65.4	60
30	3.28	.72	546	41	23	12.0	148	77.5	71
31	3.84	.78	577	44	26	13.6	174	91.1	84
32	4.50	.84	614	48	6	3.1	180	94.2	93
33	5.28	.94	659	53	5	2.6	185	96.9	96
34	6.35	1.17	720	67	0	.0	185	96.9	97
35	7.81E	1.92	803	109	6	3.1	191	100.0	98

Appendix M: Scatterplots for OCAS Scores and Number of Out-of-Class Interactions

Figure M.1

PAS Scores and Meetings during Office Hours

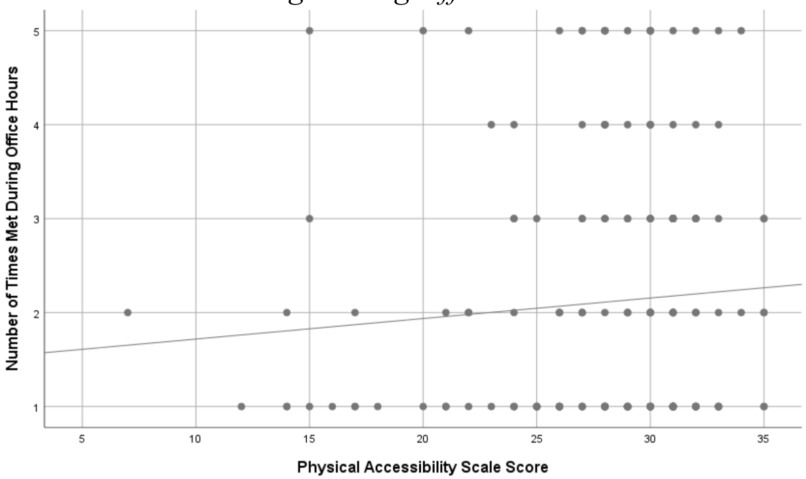


Figure M.2

SES Scores and Meetings during Office Hours

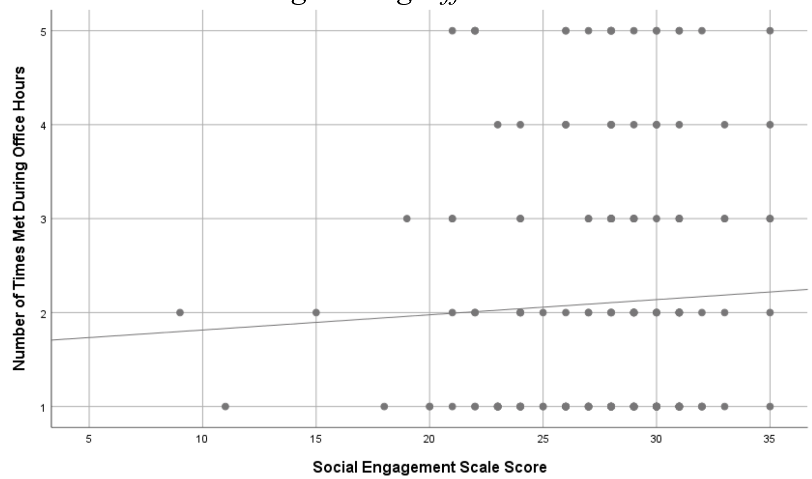


Figure M.3

PAS Scores and Meetings Outside Office Hours

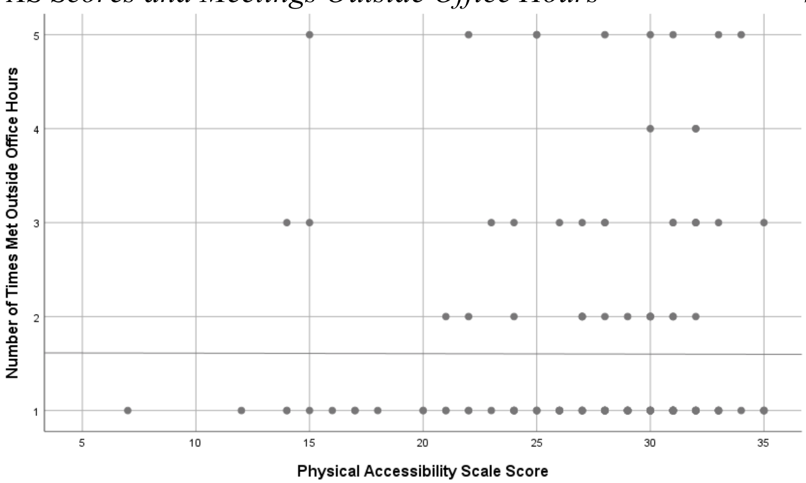


Figure M.4

SES Scores and Meetings Outside Office Hours

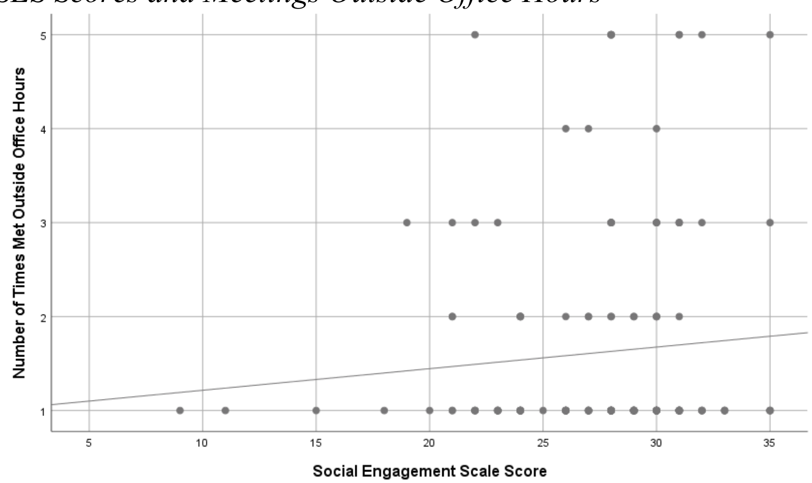


Figure M.5

PAS Scores and Informal Campus Interactions

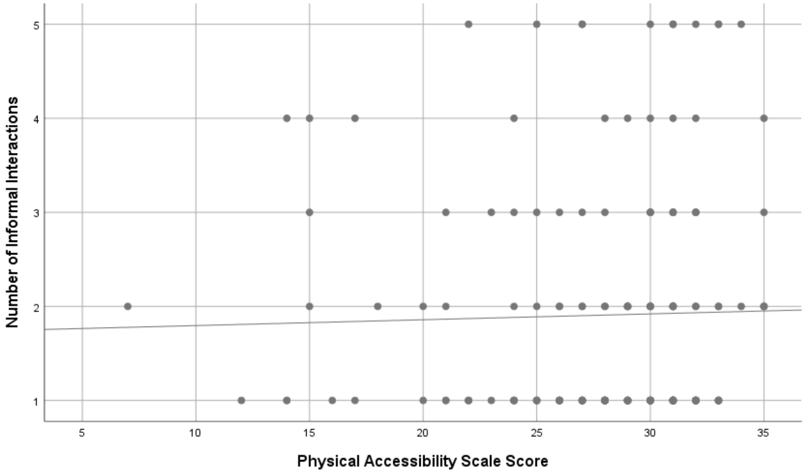


Figure M.6

SES Scores and Informal Campus Interactions

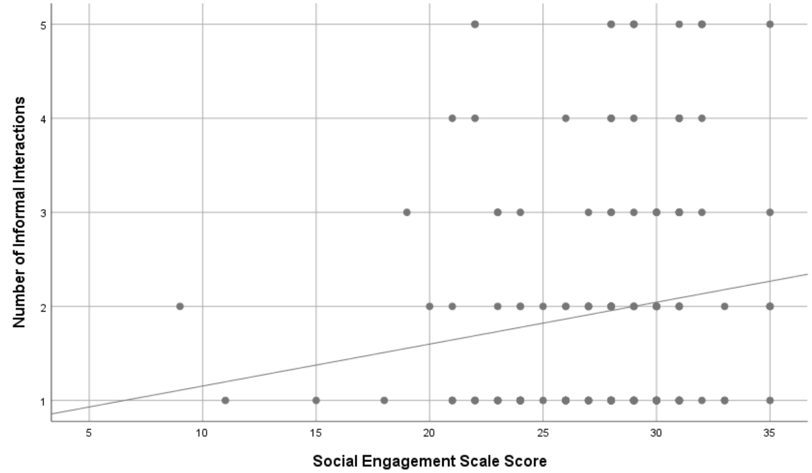


Figure M.7

PAS Scores and Email Correspondence

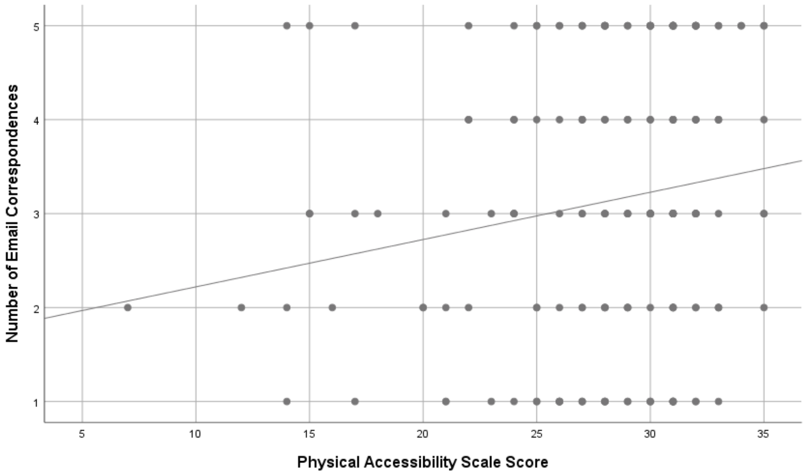


Figure M.8

SES Scores and Email Correspondence

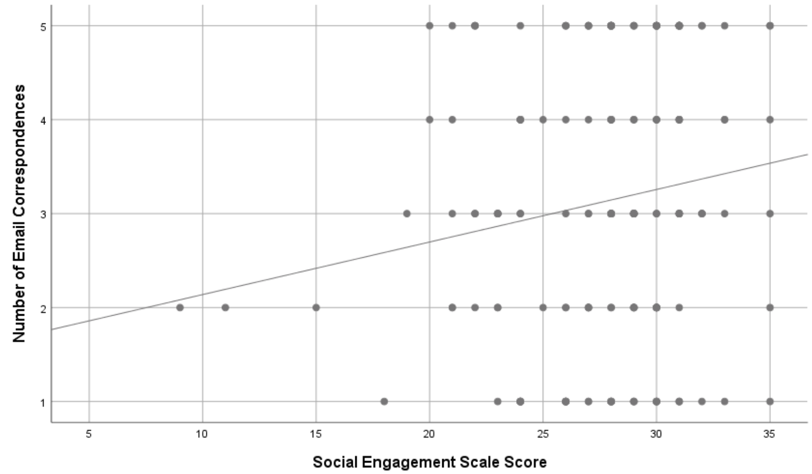


Figure M.9

PAS Scores and Interaction Before/After Class

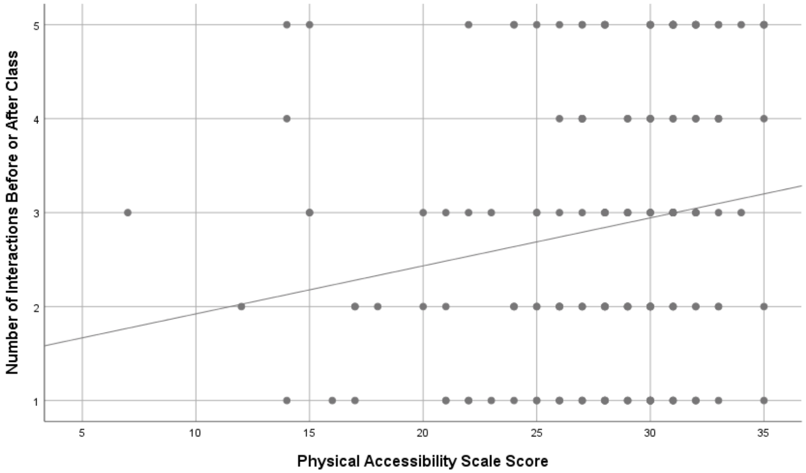
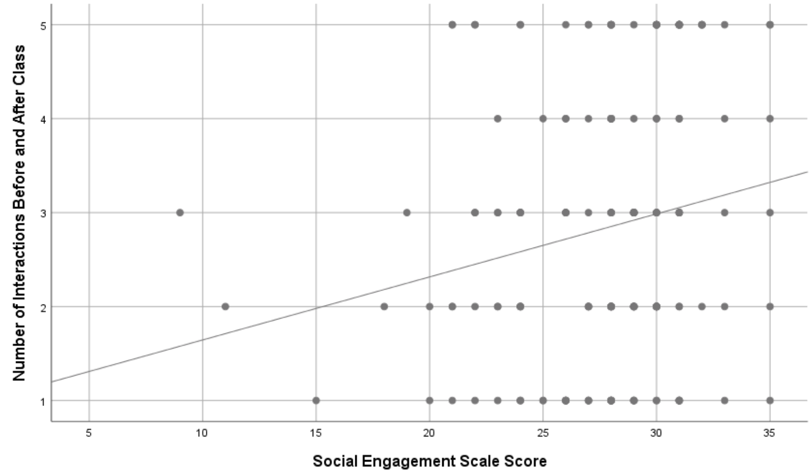


Figure M.10

SES Scores and Interaction Before/After Class



Appendix N: Logistic Regression Results for Out-of-Class Interaction Topics

Table N.1

Discussed Grades

	B (S.E.)	Sig.	Exp(B)	Pseudo-R^2	-2LL
Model 1					
Physical Accessibility Score	.004 (.030)	.906	1.004	<.001	256.186
Model 2					
Social Engagement Score	-.016 (.037)	.665	.984	.001	253.236

Table N.2

Discussed Questions about Assignment Directions or Expectations

	B (S.E.)	Sig.	Exp(B)	Pseudo-R^2	-2LL
Model 1					
Physical Accessibility Score	-.006 (.052)	.904	.994	<.001	11.054
Model 2					
Social Engagement Score	-.093 (.074)	.208	.911	.010	115.889

Table N.3

Discussed Academic Advice

	B (S.E.)	Sig.	Exp(B)	Pseudo-R^2	-2LL
Model 1					
Physical Accessibility Score	.051 (.031)	.100	1.052	.015	251.729
Model 2					
Social Engagement Score	.071 (.039)	.067	1.074	.019	247.766

Table N.4

Discussed Career Advice

	B (S.E.)	Sig.	Exp(B)	Pseudo-R^2	-2LL
Model 1					
Physical Accessibility Score	.085 (.045)	.058	1.088	.023	193.968
Model 2					
Social Engagement Score	.119 (.053)	.025	1.126	.031	193.801

Table N.5

Personal Life

	B (S.E.)	Sig.	Exp(B)	Pseudo-R^2	-2LL
Model 1					
Physical Accessibility Score	-.003 (.033)	.925	.997	<.001	223.411
Model 2					
Social Engagement Score	.088 (.046)	.056	1.092	.022	219.714

Table N.6

Discussed Letters of Recommendation

	B (S.E.)	Sig.	Exp(B)	Pseudo-R^2	-2LL
Model 1					
Physical Accessibility Score	-.057 (.054)	.291	.945	.005	82.295
Model 2					
Social Engagement Score	.006 (.076)	.935	1.006	<.001	88.443

Table N.7

Discussed Non-Class Intellectual Topics

	B (S.E.)	Sig.	Exp(B)	Pseudo-R^2	-2LL
Model 1					
Physical Accessibility Score	.049 (.038)	.199	1.050	.010	212.105
Model 2					
Social Engagement Score	.108 (.050)	.029	1.115	.029	207.256