Going Beyond the Outcome Assessment Minimum: Toward a Framework to Assess Students' Integrative Learning in a University General Education Program

Mengyun Li

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Educational Research and Evaluation

David Kniola. Chair
Stephen Biscotte
Yasuo Miyazaki
Gary E. Skaggs

October 26, 2022
Blacksburg, VA

Keywords: General Education, Interdisciplinary Program, Instrument

Going Beyond the Outcome Assessment Minimum: Toward a Framework to Assess Students' Integrative Learning in a University General Education Program

Mengyun Li

ABSTRACT

Prior research has demonstrated the efficacy of general education coursework among American college students (Ball, 2012; Rosenzweig, 2009). Traditional models of general education programs are predicated on the understanding that exposure to a broad set of educational experiences creates well-rounded graduates (Roche, 2010). However, emerging research shows the importance of integrative learning experiences including general education programs (Lowenstein, 2015). These programs are just now at the initial stages of development and implementation at colleges and universities making it possible to study direct effects on student learning. What remains, however, is limited ways to measure such learning in emerging programs. One large, research university in a mid-Atlantic state provides opportunity to construct a measure of integrated learning. This study addressed the salient literature on general education in higher education today and then used quantitative methods and qualitative methods to investigate an empirically based measure of integrative learning. Findings revealed the continuous process of integrative learning from disciplinary knowledge to application to real world and established an initial framework for assessing students learning outcomes of integration. Finally, the research provided implications for researchers and practitioners to utilize the instrument and extend it to a wider range of students and academic programs.

Going Beyond the Outcome Assessment Minimum: Toward a Framework to Assess Students' Integrative Learning in a University General Education Program

Mengyun Li

GENERAL AUDIENCE ABSTRACT

The general education framework has deep roots and is quintessential to American higher education (Ball, 2012; Rosenzweig, 2009). While general education models have evolved over time, traditional expectations are cultivating students with broad knowledge in courses that span a broad range of academic subjects (Roche, 2010). Recent research show that general education program has an impact on student development (Lowenstein, 2015). Emerging programs are beginning to shift the focus of general education from a menu-style curriculum to one that focuses on integrative learning. Currently, there is limited ways to measure and evaluate students' learning outcomes, particularly with complex learning concepts such as integration. This study was designed to develop a new measure of integrative learning. The study occurred at a large, research university in a mid-Atlantic state that implemented a novel general education program designed to focus on integrated learning in general education. The study first explored the existing literature on general education in higher education. Next, the study designed and tested a new instrument to measure integrative learning. Finally, the research provided implications for researchers and practitioners to utilize the survey. Further, it calls for additional research with different types of students across a broader range of institutions, and enhancements to the instrument that include new items to strengthen the measure.

Table of Contents

Chapter 1 Introduction 1
Statement of the Problem 4
Purpose of the Study 5
Research Questions 7
Significance of the Study 7
Significance for Practice 7
Significance for Research 8
Significance for Theory 9
Organization of the Study 9
Chapter 2 Review of Literature 11
History of General Education 11
Existing Instruments Designed to Measure General Education 18
Integration and Interdisciplinary Learning in Education 28
Integration of Knowledge 36
Integration across Disciplines 41
Integrative Learning 46
Designing Model to Measure Interdisciplinary Competence 52
Literature Gap 54
Chapter 3 Methodology 58
Instrumentation 58
Approaches to Testing 58
Towards an Initial Framework for Assessing Students' Integrative Learning 76
Sample Selection and Data Collection 93
Reasons to Combine Field Study Data 103
Qualitative Data Analysis 113
Chapter 4 Results of Study 118
Summary of Participants 118
Qualitative Data Analysis 119
Pilot study 120
Field Study 126

Results of Qualitative Data 130
Results of Quantitative Study 134
Reliability within Classical Test Theory 136
IRT Model Fit 142
Answers to the Research Questions 160
Conclusion 162
Chapter 5 Discussion 164
Summary of Findings 166
Discussion 173
Implications 174
Limitations 185
References 189
Appendix A 208
Appendix B 210
Appendix C 211
Appendix D 212
Appendix E 220
Appendix F 229
Appendix G 234

List of Figures

Figure 1 Continuum for the Interdisciplinary Curriculum Design Options..... 38

Figure 2 Process for Interdisciplinary Study of a Given Problem..... 43

Figure 3 Instrument Development Model: A Map based on Messick's Unified Theory of Validity(Purzer & Cardella)..... 74

Figure 4 Total Information Curve of Scale IWM..... 147

Figure 5 Combined Information Curve and Trait Links for Scale IWM..... 148

Figure 6 Total Information Curve for Scale IBM..... 151

Figure 7 Combined Information Curve and Trait Links for Scale IBM..... 152

Figure 8 Total Information Curve for Scale GC..... 155

Figure 9 Combined Information Curve and Trait Links for Scale GC..... 156

Figure 10 Total Information Curve for Scale ES..... 158

Figure 11 Combined Information Curve and Trait Links for Scale ES..... 159

Figure 12 The Integration of Learning Assessment Circle..... 180

List of Tables

Table 1 Comparisons between the Existing Instruments 25

Table 2 Comparison of CTT and IRT 68

Table 3 The Construct of the Instrument 80

Table 4 Scales of Measures 84

Table 5 Demographic information Item 87

Table 6 Figure Item Sample in Core Dimension 87

Table 7 Number of Items of Content by Operation in Spring 2021 88

Table 8 Participating Capstone Courses Ordered by the Number of Students in Each Course 97

Table 9 Demographic Information of Students in Pilot Study 98

Table 10 The Reliability of Different Dimensions in the Instrument 99

Table 11 Decision of the Items in the Pilot Study 99

Table 12 The Reliability of Different Dimensions in the Instrument 101

Table 13 Decision of the Items in Spring 2021 102

Table 14 Number of Items of Content by Operation in Spring 2022 103

Table 15 The Demographic Information of Participants in Spring 2021 105

Table 16 The Course Information of Participants in Spring 2021 105

Table 17 Demographic Information of Participants in Spring 2022 107

Table 18 Demographic Information of Participants in Spring 2022 107

Table 19 Descriptive Data of Common Items in Three Dimensions 108

Table 20 Independent Samples Test 108

Table 21 Independent Samples Effect Sizes 110

Table 22 Rotated Component Matrixa 111

Table 23 Open-Ended Questions in the Pilot Study 114

Table 24 Open-Ended Questions in the Field Study 115

Table 25 Examples of Coding from Excerpts to Themes 123

Table 26 Theme and Subthemes Used in Coding Process 127

Table 27 Demographic Information of Participants 135

Table 28 Construct of the Instrument 136

Table 29 CFA Fit Statistics for the Initial Framework 139

Table 30 Rotated Component Matrixa 140

Table 31 CFA Fit Statistics for the Final Framework 140

Table 32 Reliability Analysis 141

Table 33 Comparisons Across the IRT Models 143

Table 34 The Summarized Statistics of Raw Data of Scale IWM 145

Table 35 Graded Model Item Parameter Estimates for Scale IWM, logit: a(θ - b) (Back to TOC) 145

Table 36 S-X^2 Item Level Diagnostic Statistics for Scale IWM 146

Table 37 The Summarized Statistics of Scale IBM 150

Table 38 Graded Model Item Parameter Estimates for Scale IBM, logit: a(θ - b) (Back to TOC) 150

Table 39 S-X^2 Item Level Diagnostic Statistics for Scale IBM..... 151

Table 40 The Summarized Statistics of Scale GC 154

Table 41 Graded Model Item Parameter Estimates for Scale GC, logit: a(θ - b) (Back to TOC) 154

Table 42 S-X^2 Item Level Diagnostic Statistics for Scale GC..... 155

Table 43 The Summarized Statistics of the Scale ES 157

Table 44 Graded Model Item Parameter Estimates for Scale ES, logit: a(θ - b) (Back to TOC) 157

Table 45 S-X^2 Item Level Diagnostic Statistics for Scale ES 158

Chapter 1 Introduction

In a rapidly developing economy, a high-quality higher education improves benefits not only to individuals but also to the whole society (McMahon, 2009). The expectations of graduates of higher education go beyond disciplinary knowledge in an academic major to include knowledge, skills and dispositions required to be successful in the workplace and in communities. This includes an integration of abilities and skills like critical thinking and problem solving (Asyari et al., 2016; Bicer et al., 2013). To address these, colleges and universities have carefully constructed general education programs and curriculum designed to produce "well rounded" graduates. These programs have evolved over time and continue to be shaped by both emerging research on learning and societal needs.

Different from many other education systems worldwide, the American model of general education formed its roots in the nineteenth century, it later changed to a more prescriptive model in the 1920s to better align with workforce needs (Bisesi, 1982). These modern changes were largely due to the great depression and growing purposes of higher education for workforce development as opposed to the purposes of free thinking and enlightenment in prior eras.

After nearly 100 years development, most undergraduate students are now required to take general education courses in addition to their major field of study. Often referred to as "general education requirements" or "common core curriculum," these courses continue a tradition of holistic education. Although most institutions and instructors understand the broad goals of such general education curricula, measurable

learning outcomes of general education has been elusive (AAC&U, 2007). In fact Cohen and Brawer (2008) claim, “General education has remained a noble idea but a practical backwater in most of American higher education” (p.374). This observation underscores a critical need to find objective standards to assess the benefit of general education as it relates to important learning outcomes and to build a case for both academic and public confidence in general education (Seifert et al., 2008). In addition to this, the majority of studies to date focus on narrow academic outcomes; however, relatively few empirical studies have tried to identify effects on the social dimension of learning and development (King et al., 2007).

For the purposes of this study, a single institution will be selected to examine the development and efficacy of an instrument designed to directly measure a central learning outcome of general education. This is markedly different than the rubrics and processes proposed by others to assess general education learning outcomes. For example, The American Association of Colleges and Universities which is a global membership organization devoted to progressing undergraduate liberal education is now initiating VALUE as an authentic approach to measure the skills, abilities, and dispositions in liberal arts education (AAC&U, 2022). The 16 VALUE rubrics they developed provided a framework to assess students’ cumulative progress in liberal arts education, it mentions integrative learning, but it lays out into the language of individual courses, disciplines, and campuses, not regard liberal arts program as a whole context. It mainly focuses on learning outcomes such as critical thinking, integrative learning is only one category of the 16 rubrics, it is not the main goal of the assessment, like many other

assessment tools used in general education. Other assessment instruments focused on the particular learning outcomes and most of them take integrative learning as only one dimension as the learning outcomes.

The university that the research implemented fulfils its role as a land-grant institution by fostering a collaborative environment that integrates technology into all disciplines, it is a Research Intensive (Research I) university located in Virginia. It offers about 280 undergraduate and graduate degree programs to more than 37,000 undergraduate, graduate, and professional students across the commonwealth(Virginia tech, n.d.) The university provides undergraduate students a general education program, one that helps them to integrate their learning across disciplines and tackle challenges of the future, and minor program in tis an essential component of the new general education program. The program is thematic interdisciplinary program that are built such that students can both meet general education requirements (e.g. natural sciences, humanities) while also diving deeper into a broad real-world topic. An interdisciplinary program is only benefit for learners in a complex system set, if a phenomenon is not generated by a system which complex interdisciplinary study is not necessary (Newell, 2001). The topics sustainability, innovation, global food security, and social justice in minors are popular topics that need student to dive into from a complex context, in this way, the minor program is a perfect approach for students broaden a complex and global horizon. There are 27 Minor programs, and to-date, over 200 students have graduated with a Minor program and over 750 students are currently enrolled in one (All data comes from the official website). Each topic includes at least 3 core concepts, all students undertaking

the minor program must meet the requirements of at least 18 credits with a minimum of 6 at the 3000-4000 level course. Students in these minors involve in foundational learning, exploration from different viewpoints, and then bring it all together in an applied, project-based capstone course, they have an opportunity to master a well-rounded impression over the theme instead of solely fragmented knowledge based on an interdisciplinary learning experience. Beyond this, the capstone courses also provide student a chance to have cooperative learning with the peers who own different backgrounds. Students will have more social engagement than the solo course in a traditional general education program.

Statement of the Problem

Although there are numerous tools and methods for measuring general education outcomes and programs (Dee Fink, 2009; Greene, 2003; Marshall et al., 2005; Yu et al., 2019), there is scant literature on measures of learning outcomes emphasizing integrative learning in an interdisciplinary general education program. Given the rising importance of progressive approaches to general education, new research is needed to substantiate the shift from traditional general education to new models. The problem becomes more complex as specific institutions move towards new, more sophisticated general education programs which include components like thematic clusters and integrated minors.

The program of focus for this study uses an innovative model of general education which is integrated and interdisciplinary. The general education program is progressive because the program is thematic, cross-disciplinary programs that allows students to examine important topics from a variety of perspectives while completing

general education requirements along the way. The format of the program is very innovative across the country and the number of students in the program is continuing to grow. The purpose of this thematic format is to teach student integrative learning, but whether the program achieve its priority target of integration remains unknown to the program evaluation, and the mechanism to measure students' experiences in this progressive program is also an open question among academic leaders and administrators. Developing a measure for the program is a perfect topic to build a stone mile for assessment practice of interdisciplinary program. In addition to this, both curriculum construction and the administrative environment of the program provided many supports to set up this research, because all stakeholders desire to know the learning outcomes of integration through integrative learning.

Perhaps more importantly, instructors and students seek clarification and a deeper understanding of the meaning of their own activities and engagement. Here, "efficacy" is to know why this program is a better educational choice than isolated courses, to inform strategies that improve the learning experiences, and to focus the educational intentions of various stakeholders. Practical questions remain such as "Does the minor general education program really promote student integrative learning?", "Does it really help students to integrate knowledge?", and "Does it help students to master the ability to integrate for their future life?"

Purpose of the Study

This study proposes a new instrument that will address the specific interest of integration. It proposes a measurable student learning outcome called integration that is

central to interdisciplinary general education programs. The results of this study will provide a critical first step to measure a new learning outcome and inform assessment practices relative to integration.

Interdisciplinary general education programs have not yet been studied widely because they remain nascent in American higher education. While interdisciplinary academic programs have gained popularity, its use in general education is new. As such, measuring learning outcomes in this educational space, those that are unique to "interdisciplinary course learning" remain understudied.

The main purpose of my dissertation will be to fill this gap by providing an empirical measure of integration. Developing such an instrument will allow a range of stakeholders' insight to students' perceptions of their integrative learning experiences and outcomes from the progressive general education program that are interdisciplinary. This study will investigate the development of an instrument designed to measure integration with the expectation that its use will be generalized to other interdisciplinary general education programs.

To fulfill these purposes, my dissertation will be organized into several steps that, when considered collectively, will advance our understanding of student integrative learning experiences and learning outcomes from an interdisciplinary general education program. In its entirety, my study aims to trace a history of general education that informs today's emerging models. The research will present prior research and practical experiences to explore the general goals of today's general education. Then it will dive into the dimensions of learning outcomes and learning experiences in an interdisciplinary

program, to analyze what factors are most essential to affect students' integrative learning. Next, the research will collect and analyze data at the course level to examine psychometric properties that inform the development of a new instrument. Part of this study will also involve a faculty survey to better understand the instrument.

Research Questions

The main purpose of the study is to develop an instrument to investigate whether the program promotes students integrative learning in the program. The research questions for this dissertation are:

1. Can we develop an instrument to measure a student learning outcome which is called integrative learning in a thematic general education program?
2. Is there a measurement model that best fits the instrument?

Significance of the Study

This study has important implications for educational practice, research, and theory. The following sections provide anticipated significance that may enhance understanding the application of interdisciplinary to general education, it set up a foundation of assessment to the similar programs in U.S. higher education institutions.

Significance for Practice

This study was significant for several campus constituencies. One group that might benefit from the results are working in the administration departments of undergraduate education. These university policymakers implement academic policies to support and enhance the quality of educational programs. Similarly, those charged with oversight of general education programs have interest in empirical evidence of student

learning. The instrument is designed to measure important aspects of student experiences, learning outcomes and the program satisfaction. There remains wide interest within institutions to evaluate whether general education programs meet stated goals, particularly those that are interdisciplinary. Informed decisions remain at the forefront to improve program quality, promote instructor and student engagement, and understanding experiences among minoritized student groups.

Further, this study establishes a basis for best practices to measure important student learning outcomes through assessment. Lessons learned in this study may inform assessment professionals, particularly those concerned with general education, in how to construct similar instruments or adapt this one to unique circumstances.

Significance for Research

The present study will have significance for future research on student learning outcomes assessment and general measurement in education. Future studies might continue work started here by collecting a longitudinal data to not only examine the impact of integration, but to also improve the quality of the proposed instrument. If a future study can accumulate large scale data, researchers can do the analysis of quality of integrative learning experiences and outcomes differences among students, including differences by gender, race, and other important characteristics of learners. These are also good way to know more about the students' integrative learning outcomes in a general education program from an interdisciplinary perspective. The data collected by the instrument give an opportunity to improve the existing assessment projects of the current

general education programs, but also for other similar interdisciplinary programs in the future.

Significance for Theory

The study will be significant in terms of future theory that informs measuring student learning. To date, most assessment projects of student learning experiences and learning outcomes were only suitable for the student's whole experiences during the undergraduate studies and remain specific and pragmatic to a single institution. A few empirical research are mainly limited to an interdisciplinary program or a general education program. This study strictly followed the rules of social science, finally developed a new framework in the instrument which will offer an insight into the combination of both interdisciplinary programs and general education programs. It might be used to expand existing theory on the student learning experiences and such novel learning outcomes in the future.

Organization of the Study

The present study is organized in five chapters. Chapter One introduced the topic of the study, the research questions, the purpose of study, and the significance of the study. It presented why I am interested in doing this research, and the reasons why this research is worthwhile. The second chapter reviewed the literature relevant to the study, this process helped to figure out the factors to build the initial framework of the instrument for pilot study and second wave of data collection. In my plan, the chapter Three will describe the methodology of the study. It includes the construct map, theoretical framework, internal construct, external model, measure scales and other

essential information of the instrument. Then it introduces the sampling techniques, the procedures used to collect and analyze the data. The process of analyzing data was different from conventional quantitative research, it mainly did item analysis to show the discrimination and difficulty of each item, then tried to use different models to fit the instrument. The fourth chapter describes the results of the study, it explained and compared the results in different models, made the decision to revise items. The final chapter discusses conclusions, the final developed instrument, limitation and the implications for future practice, research, and theory.

Chapter 2 Review of Literature

This chapter will systematically discuss previous literature related to the research questions. As context to build a framework for the survey, first I summarized the substantive detail on the work of general education, existing student learning outcomes assessment instrument, these will help me to find some clues about some essential learning outcomes of general education and will be strong support for the process to construct the instrument. Then I will document the literature on the interdisciplinary programs, this part includes the interdisciplinary curriculum development, knowledge integration, and high skilled integration, highlight the important characteristics and learning outcomes from the interdisciplinary program. From this part, we concluded some of the indicators that could be used in the instrument. Finally, I went into student engagement and active learning. All this work will be the clues and references to build the framework for the instrument.

History of General Education

General education courses were widely acknowledged to provide a groundwork for further developing the skills, competencies, and dispositions which were the essential learning outcomes in higher education (Nelson et al., 2009). When thinking about what the learning outcomes of new and emerging general education models are, it is informative to first examine the history of such programs in the US higher education.

In the dawn of higher education in colonial America, general education and liberal education shared the same meaning. It is widely recognized that Harvard college, established in 1636, constructed a list of common core knowledge and skills for liberal

arts. Harvard College followed a set curriculum that consistently emphasized rhetorical principles, rote learning, and constant drilling during its first two hundred years (Harvard College Handbook for Students, 2020-2021). Eventually a curriculum that increasingly emphasized professional knowledge moved in to challenge historical ideas of liberal arts education across academic disciplines and courses (O'Banion, 2016). Until then, the modern perception of general education started early in the nineteenth century. Thomas Jefferson was a leading advocate of establishing a society in which the educated citizen had a constitutive role as an involved representative of knowledge and "know-how" (Holowchak, 2014). The president of Harvard, Charles William Eliot demanded a place for the professional knowledge as well as the humanities in the program of liberal education. Over time, the idea of a common core evolved into a form of general education, attempting to unify core knowledge and skills learned by undergraduate students (O'Banion, 2016).

Beginning in 1919, Columbia University played an early important role in general education by creating a special course list. Under the title "War Issues to Contemporary Civilization" it was a mandatory overview of knowledge and resources designed to help students understand the broader world and their role in it (O'Banion, 2016). While Columbia provided inspiration to others, the University of Chicago constructed a new plan with ambitions to be "the most thoroughgoing experiment in general education of any college in the United States" (D. Bell, 1966, p. 26). Although the course offered by this program was meaningful in history, the learning goals of the program was fostering

gentleman for the upper classes based on the ideal standards in 19th century (Stevens, 2001).

A critical turn happened early in the 1930s when Robert Hutchins, while working as president at Chicago, began a remarkable reformation of the liberal arts. Hutchins strongly advocated for a curriculum to prepare student for life through democratic education. (Stevens, 2001). An educational philosophy developed that aligned education, teaching, knowledge, and perennialism in education.

By the 1940s, a general education movement was in progress in American higher education. The Harvard “Redbook” study(Wilson, 1946) stated “…a quest for a concept of general education that would have validity for the free society which we cherish" (p. xiii). It regards general education as "liberal education” and specifies a core curriculum of six required courses(Wilson, 1946, p. 196). This is the first evidence of a prescribed curriculum. With the rapid development of career-oriented education, college and university faculty created professional courses to attract more students. This increasing specialization within academic disciplines caused the fragmentation of knowledge in the university later and created philosophical tensions between specialized and liberal education.

In the decade that followed, a belief in the universality of human nature intersected with the general education movement. As stated in Higher Learning in America: "One purpose of education is to draw out the elements of our common human nature, these elements are the same in any time or place” (Wilson, 1946, p. 66). In the early 1950s, Johnson (DATE) listed 12 competency goals that should be showed in a

person who received a general education. Many of these remain evident in learning outcomes today. This work advanced the idea that the aim of general education was to develop the whole-minded students by re-building the balance between general and specialized knowledge. This list was the result of analysis of the catalogs that included general education among hundreds of community colleges across the nation throughout the 1950s and 1960s (B. L. Johnson, 1952, p. 21,22). The list showed that after the general education movement, the main goals of general education transferred to human development, including individual and social adjustment which contributed to a happy life.

The resulting list of critical learning outcomes inspired the general education in the decades that followed. In early 1980s, Hursh (1983) and other researchers presented strong evidence that general education helped student develop a capacity for discovery and exploration of thinking, inquiry, and searching for patterns of meaning that were embedded in the disciplines. Following this, the Carnegie Report of 1987 spoke of general education as an "integrated core" of common learning. The “integrated core” conception came from liberal arts education.

Although general education and liberal arts education had some overlap, they were different in ideological orientations, pedagogical methods, curricular structures, and ultimate aims (Erickson, 1992). Erickson argued that liberal education is based on “existence of an independent reality consisting of fixed laws and absolute truths” while general education is grounded in “instrumentalism's subjectivist assumption of the psycho-physical interdependent existence of reality.” Due to their different philosophical

foundations, liberal education focused on timeless knowledge as well as timely social issues such as civil rights and gender equality (Bryant, 2003; Weakliem, 2002). Further, it tried to teach students constant truths of the past as the model by which students can still be guided in present days (Bloom, 2008; Van Doren, 1959). On the other hand, general education placed strict emphasis on the needs of the students and their future career.

Today, general education is informed by a more complex and modern context related to information technology, diversity awareness, and global understanding. With the explosion of knowledge resulting from globalization and the Internet, general education is often not seen as an interconnected core of courses required of all students. It largely became a set of courses loosely connected to core disciplines from which students make enrollment decisions in two or three courses from among a feast of a hundred or more choices (O'Banion, 2016). With this new background, seemingly the goal was to combine the best from both liberal education, that focused on broad knowledge, and career-orientated education, that emphasized specialized knowledge, into a new format for general education that would offer students a range of course choice designed to improve their learning experience and lead to better jobs after undergraduate studies.

This new format of general education prepared students for the labor force through in-depth study in specific fields while at the same time emphasizing academic breadth by having a good perception of various fields of knowledge. During this era, learning outcomes reflected this blend of liberal arts and career-oriented education as

critical thinking, problem-solving, communication, and collaboration became tantamount to education. (O'Banion, 2016).

This approach to general education is substantiated by regional accrediting bodies that are charged by the U.S. Department of Education to oversee quality assurance. The six regional accrediting agencies define general education quit similarly and address basic learning outcomes. The listed learning outcomes from accreditation agencies are based on program level learning outcomes and are seen as an essential part to guarantee the quality of general education program. While liberal education, which was to help students to become informed citizens, general education in fact, uniquely fostered a broad range of empirically substantiated good practices in undergraduate education. General education made students well-rounded graduates that were flexible and adaptable to the changes of the world(J. Fredericks Volkwein, 2009). The required program-level learning outcomes of six regional accrediting agencies focused on breath of knowledge, skills for global economy and lifelong learning. To more specific, students who finished the credits in general education need to identify and multidisciplinary research question, to complete projects that integrates what they learned in their courses to a real research question (Greene, 2003). At the same time, students were required learnt to explore the interrelationship of knowledge in the increasingly global society, to examine the people, values, and societies of different societies, to develop the global perspective to adapt to an ever-changing international environment (Dee Fink, 2009). From this perspective, the focus of learning outcomes from general education is changing according to the context.

We need to investigate the ability to integrate knowledge with different experiences in general education.

In today's US higher education system, Undergraduate Institutions and Programs contains information reported to the U.S. Department of Education directly by recognized accrediting agencies and state approval agencies (Hegji, 2017). The six regional accrediting agencies define general education in the different ways, it concludes the basic program-level learning outcomes of written communications, reasoning in both social and scientific areas, humanities, ability to become a better learner and citizen.

The listed learning outcomes from accreditation agencies were based on program level learning outcomes. From their official documents, all standards from six religions emphasized the breath of the knowledge, some agencies highlighted the integration in the learning process across disciplines, it focused on breath of knowledge, skills for global economy and lifelong learning which were an essential part to guarantee the quality of general education program from the governmental accountabilities. While liberal education, which was to help students to become wise citizens, general education in fact, uniquely fostered a broad range of empirically substantiated good practices in undergraduate education. General education made students well-rounded graduates that were flexible and adaptable to the changes of the world (J. Fredericks Volkwein, 2009). To more specific, students who finished the credits in general education need to identify and multidisciplinary research question, to complete projects that integrates what they learned in their courses to a real research question (Greene, 2003). At the same time, students were required learnt to explore the interrelationship of knowledge in the

increasingly global society, to examine the people, values, and societies of different societies, to develop the global perspective to adapt to an ever-changing international environment (Dee Fink, 2009). From this perspective, the focus of learning outcomes from general education is changing according to the context. We need to investigate the ability to integrate knowledge with different experiences in general education.

Existing Instruments Designed to Measure General Education

With this understanding of the history and trajectory of general education, now turns the attention to the literature on empirical studies designed to measure such programs and their intended learning outcomes. Attention to assess students' experiences and learning in higher education has been powered by the recognition that insights into student development and learning can inform pedagogy as well as program-level improvement (Astin, 1985). Assessing students experiences and learning outcomes is an essential activity that allows someone to judge whether a program is working, and also vital for improving the quality of higher education (Steedle, 2012; Association of American Colleges and Universities and Council for Higher Education Accreditation 2008). The quality requirements in the form of federal, regional accreditation, together with an institutional self-commitment to improve, were the primary motivations of assessment. Stated learning outcomes within academic programs is less a rule than a norm nowadays. Whether perceived as a mandate or integral for improvement, assessment has become a widespread practice in American higher education. By 2013 more assessment projects were being conducted than the past with nearly 84% of all colleges and universities had stated learning outcomes for their undergraduates, a growth

of 10% from 2009 (Kuh et al., 2014). The Federal government established its own system to assess education efficiency. One of the most widely used example of this is the Integrated Postsecondary Education Data System (IPEDS). The IPEDS system, as well as related surveys, is conducted annually by the U.S. Department of Education's National Center for Education Statistics (NCES). IPEDS gathers data from every college, university, and "technical and vocational institution that participates in the federal student financial aid programs" (Integrated Postsecondary Education Data System, 2021). While overarching in nature rather than specific to learning, it remains a key source of information to reliable quality.

General education is an essential part of undergraduate education, educators and researchers were discovering to broaden assessment approaches to enhance both teaching and learning, by the mid-1980s national interest started to pay attention to the state of general education in the United States institutions of higher education (Yeilding, 1987). States have utilized various assessment tools or asked institutions to implement the tool for different purposes and with variable levels of achievement in general education. These efforts have resulted in a series of assessment methods and of differing impacts on institutions (Champney & Edleman, 2010). For example, Virginia required public institutions to select an approach to assess general education targets and submit results as a statewide accountability variable while South Dakota, mandated all students to complete a standardized examination of general outcomes such as reading, writing, math and critical thinking before graduating. Furthermore, Florida required that all students reach specific scores on a standardized examination as a pre-requirement for advancing

from their second to third year during their undergraduate study (Kuh & Ewell, 2010). The institutional level assessment provides more information for decision-support for the whole university while program-level will mainly for accountability for different stakeholders. Most institutions use a combination of institution-level and program-level assessment approaches (Kuh & Ikenberry, 2009). The National Survey of Student Engagement (NSSE) is widely used instrument by institutions to assess student's learning outcomes at an institutional level. It is a large, standardized survey project in the USA, Canada, it collected information of four-year colleges and universities about the first year and senior students' participation in programs and activities that institutions provide for their learning and personal development (NSSE, 2020). The survey was based on the Astin's I-E-O model and other student engagement theories, questions in NSSE survey included students' experiences in their whole undergraduate program. Although this national survey was not developed specially for general education, it emphasized higher-order learning, collaborative learning, reflective and integrative learning, which provided inspirations for developing assessment instruments especially for the theme-based general education. In addition to these important scales, the approach of using student perception of experiences to predict student outcome was also adopted by other assessment instruments. ''Course experience'' was a central element of assessment that course experience is a summary of students' perceptions about the quality and usefulness of instruction (Volkwein, 2009). The vital factors that can affect student awareness of toward instructional quality are course goals and expectations, and course workload in assessment of learning experiences (Budd et al., 2013; Libarkin, 2001). The stakeholders

use classroom assessment as a reference approach to improve the quality of the program for general education. The CEQ was developed by Ramsden (Ramsden, 1991; Wilson et al., 1997) as a teaching assessment indicator, focusing on the classroom environment of teaching which had found were linked to approaches at different levels to learning, and higher quality learning (Chalmers, 2007). This instrument provided a valuable indicator about student perception of satisfaction regarding the courses that they study for the later assessment. The scales in CEQ include Good Teaching, Clear Goals and Standards, Appropriate Assessment and Appropriate Workload (Chalmers, 2007). Different from NSSE which was mostly emphasized the experiences during the four years undergraduate study, the instrument itself included an outcome scale of Generic Skills, and an Overall Satisfaction at the course or program level (Talukdar et al., 2013). With the development of different testing agents and online testing, to participate in standardized tests (e.g., ETS Proficiency Profile, CAAP, CLA), or postsecondary attainment testing programs has been another mainstream way to assess student learning outcomes in general education (Steedle, 2012). These tests reflected student gains of higher-order learning outcomes, indicated relative strengths and limitations in general education programs. The ETS Proficiency Profile is a single convenient assessed four core skill areas including reading, writing, mathematics and critical thinking that the Voluntary System of Accountability (VSA) has chosen as a standard of general education outcomes (ETS® proficiency profile, 2021). The Voluntary System of Accountability utilized value-added method to generate value-added scores for the which is to evaluate institutional core educational outcomes from general education (e.g., critical thinking, written or communication). The

Collegiate Learning Assessment (CLA+), developed by the Council for Aid to Education (CAE), assessed the college and career readiness competencies of critical thinking, problem solving, and effective written communication, providing students and educators with valid and reliable diagnose for students' strengths and help them to prepare for next step in career path. These are the skills most in demand by both higher education and employees, are also important learning outcomes from general education. The Valid Assessment of Learning in Undergraduate Education (VALUE) (AAC&U, 2009) was a defined competency from general education theoretically, considering Essential Learning Outcomes for the twenty-first century (LEAP) which are Inquiry and Analysis, Critical Thinking, Creative Thinking, Written Communication, Oral Communication, Quantitative Literacy, Information Literacy, Reading, Teamwork, Problem Solving, Civic Knowledge and Engagement. AAC&U's Integrative Learning VALUE Rubric (henceforth "IL Rubric") was also used to foster student's integrate learning ability. To explore how students' perception of integrate learning across different learning environments, Wade (2021) used the "IL Rubric" as basic to develop the Integrative Learning Survey which measured five steps in total including reflection and self-assessment, connections to experience, connections to discipline and transfer knowledge, the whole process to learn. Also, a number of assessment projects were proposed by discipline because different curriculum setting in general education also has different educational goals. For example, science curriculum enables student learning of ''broad knowledge of human cultures and the natural and physical world, including social sciences, science and mathematics, humanities, histories, and the arts'' as well as

‘‘intellectual and practical skills, including effective writing, inquiry, quantitative and information literacy, and teamwork and problem solving’’ (Laird et al., 2009). ACS was an instrument developed for course experience in a general education integrated science course in a big southwestern public university (Libarkin, 2001). The requirement of the new science course structures was emphasizing on the processes of science rather than the traditional methods focusing solely on knowledge attainment (Cashin & Downey, 1995; Smart & Ethington, 1995). The change of the course requirement was meaningful to the paradigm of general education, it tried to explore the interconnections within the same discipline. Although the assessment instrument was specially developed for the science course in general education, but we could clearly see, with the development of general education itself, the purpose started to change from added value of detailed competency to the ability of learning and application. Table 1 provides a summarized comparison of existing and widely used instruments. This table compares various instruments that are intended to measure learning outcomes of general education, each having a specific focus. This collection from prevailing literature clearly identifies notable gaps, specifically the nature of these instruments points to general experiences rather than targeted learning outcomes. Importantly, while these instruments provide a wide-angle snapshot, each provides insight to development of a new instrument to measure course-level learning experiences.

Although there were so many research and assessment project on student learning outcomes, most of them focus on specific skills such as writing or thinking, because most of the courses in traditional general education were separate. Some instruments tried to

find connections within academic discipline. The assessment conducted by Prince George's Community College (PGCC) integrated assessment of course, program, and general education. PGCC's used "All-in-One" method, allowed faculty to depict students' unconnected skills by using rubrics to assess and grade key projects across program curricula and then entering the data into a centralized database, this made the whole process, including grading and the evaluation of course, program, and general education learning outcomes, were assessed in a single assessment system, the "All-in-One" approach did not treat the course in general education as individualized, separated courses, it tried to address the learning outcomes from general education throughout the program (Richman & Ariovich, 2013). The assessment system was innovatively tried to find and create interconnections within general education. University of Massachusetts Amherst developed a learning mapping tool for the integrative learning in general education. It contains two elements to the analysis, the first element is a quantitative

Table 1

Comparisons between the Existing Instruments

Instrument Name	Affiliation	Subscales	Target groups	Subject Area or Model	Focus
ETS Proficiency Profile	ETS	Mathematics; Writing; Reading	College students	General education outcomes from the Voluntary System of Accountability (VSA).	Reading, Writing, Mathematics and Critical Thinking
NSSE	Center for Postsecondary Research (CPR) in the Indiana University School of Education	Academic Advising; Career & Workforce Preparation; Civic Engagement; Development of Transferable Skills (Updated in 2021); Experiences with Online Learning (New in 2021); Experiences with Writing; First-Year Experiences and Senior Transitions (Updated in 2021); Global Learning HIP Quality (New in 2022) Inclusiveness and Engagement with Cultural Diversity	First-year and Senior Students	Tinto's I-E-O Model; Student Engagement Theories	Student Learning and Development Institutional Research

CEQ	Centre for the Study of Higher Education (CSHE) and the Assessment Research Centre (ARC) of The University of Melbourne.	Good Teaching. Clear Goals and Standards Generic Skills Appropriate Assessment. Appropriate Workload Emphasis on Independence	Undergraduate Students	Business, commerce, computing science etc.	General Skills and Overall Course Satisfaction
Collegiate Learning Assessment (CLA)	Council for Aid to Education (CAE)	Data Literacy; Critical Reading and Evaluation; Critiquing and Argument; Writing Effectiveness; Writing Mechanics	College Students	Value-added Model	Critical thinking, Analytic Reasoning, Real-world Problem solving, and Written Communication Skills
Valid Assessment of Learning in Undergraduate Education (VALUE)	AAC&U	Knowledge of Human Cultures and the Physical and Natural World; Intellectual and Practical Skills; Integrative and Applied Learning;	College Students	Essential Learning Outcomes for the Twenty-first Century (LEAP). Voluntary System of Accountability	Inquiry and Analysis, Critical Thinking, Creative Thinking, Written Communication, Oral Communication, Quantitative Literacy, Information Literacy, Reading, Teamwork, Problem Solving, Civic Knowledge and Engagement

Assessment of the General Education Program	A Northeastern Public University	Content. Critical Thinking; Communication; Connections	College Students	General Education	Their college experiences and subsequent professional training; Their careers and productive lives; Community engagement and informed citizenship; A diverse and rapidly changing world; A lifetime of learning
Attitudes and Conceptions in Science (ACS)	A Southwestern Public University	Attitude toward Learning Science; Attitude toward Science; Conception of Science	College Students	General education course in geology or planetary science, introductory geology	"Conception of the nature of science, as well as attitudes towards science and learning science

translation of student-identified connections, the second is to uses inductive approach in qualitative method to code students' descriptions of the kinds of connections and the impact of those connections(Stassen, 2019). This assessment project started to consider the integrative learning in general education. But it did not generate a consistent instrument that can be applied to other general education programs.

The existing instruments provide many insights for the new instrument development, from assessment approaches to items development. But few instruments focused on the how integrative learning in the connected courses across different disciplines affect students, and the existing research and instruments were seldomly regard the courses in one general education as continuous entity or connected courses.

Integration and Interdisciplinary Learning in Education

Speaking of interdisciplinary, there were two other similar words multidisciplinarity and transdisciplinarity. Multidisciplinarity was less integrative, it often referred an impermanent or weak combination of elements from multiple disciplines (Chubin, Porter, Rossini, & Connolly, 1986; Committee on Facilitating Interdisciplinary Research, 2004), while another synonymous word transdisciplinarity focused more on constructing theories that surpassed the boundary of disciplines (Lattuca, 2001). The term interdisciplinary drew on disciplinary perspectives integrating their insights through construction of a more inclusive perspective(Klein, & Newell, 1997). 'Inter' is a prefix can mean establishing a interaction between and joining together, as in 'international', and at the same time separating and keeping apart (McQuillan et al., 2022, pp. 104–113). 'Interdisciplinary' can indicate making

connections across the different disciplines; but it can also mean creating a sort of undisciplined area trying to surpass disciplinary boundaries. Interdisciplinarity means any form of dialogue or interaction between two or more disciplines(Moran, 2010, p. 14). Klein and Newell (1997) proposed a survey of contemporary interdisciplinary studies for the Handbook of the Undergraduate Curriculum, it included diverse motivations for interdisciplinary study, general education is inside. They defined interdisciplinary studies in the following terms: "Interdisciplinary studies may be defined as a process of answering a question, solving a problem, or addressing a topic that is too broad or complex to be dealt with adequately by a single discipline or profession . . . is draws on disciplinary perspectives and integrates their insights through construction of a more comprehensive perspective" (pp.393-394). Julie Thompson Klein (1990) offered a general definition of interdisciplinarity: “Interdisciplinarity is a means of solving problems and answering questions that cannot be satisfactorily addressed using single methods or approaches” (p. 196). “Interdisciplinary” studies was a subset of integrative learning that fosters connections among disciplines and interdisciplinary fields(Klein, 2005).

As a transformative President of Harvard University, James B. Conant pointed out undergraduate education had become too specialized to educate youth to know themselves and explore the world in which they lived. Under this concern, he and the Harvard committee wrote Education for a Free Society, the so-called Redbook, consequently proposed "broader, more inclusive, thematic sets of courses in the various major divisions of knowledge”. It pointed out thematic courses was a good approach to

break disciplined borders. The needs to integrate the intellectual, social and emotional dimensions of undergraduate student learning in higher education had been voiced regularly throughout the last decade(American Council on education, 1949).Interdisciplinarity became important for more than one reason and addressed more than one type of problem. Moreover, the need for integrative problem solving and team work has been a key motivation in the professions, in industry, and in government(Klein, 2001). Interdisciplinary approaches were also necessary for confronting the most critical technological and socio-technological challenges facing the world (Hartesveldt & Giordan, 2008; Omplication et al., 2006) On the account of new fast-changing society, providing student endure abilities to meet the increasing requirement of the challenges became the center of interdisciplinary program.

The booming number of interdisciplinary programs was solution under this context, to improve the quality of these programs, developing operational definitions of interdisciplinarity through learning outcomes was the essential first step to developing and assessing the effectiveness of interdisciplinary programs. After the development in the past decades, many educators had realized its benefit from interdisciplinary perspectives. In early 80s, by combining learning outcomes from more than one discipline, students got the chances to challenge conclusions separately and eventually work toward a more comprehensive understanding of the problem, the "process" of challenging have to occur. (Hursh et al., 1983). In early 90s, Newell(1990) concluded that interdisciplinary program could brought students “more sensitivity to ethical issues, ability to synthesize or integrate, enlarged perspectives or horizons, more creative,

original, or unconventional thinking, more humility or listening skills, sensitivity to bias". In introducing of Interdisciplinary Undergraduate Programs: A Directory(Edwards, 1996), William Newell argued that interdisciplinary education was moving from the "radical fringe to the liberal mainstream". Institutions increasingly offered "interdisciplinary" programs as markers of their competencies to produce a new generation of intellectuals and professionals (Lattuca et al., 2004).

Problems exist in today's much more complex, globalized society context. Interdisciplinary took on a meaning that includes a broader range of stakeholders, including practitioners and the public in its attention on solving real problems(Michael Gibbons et al., 1994; Klein, 2005). This makes each of the problems or issues as a particular complex system, help interdisciplinarians better understand such phenomena or problem, continuously and collectively, which is the nature and conduct of interdisciplinarity. Since because of the complexity of the interdisciplinarity, although instructor had good intentions when they planned to teach in interdisciplinary courses, when they faced the question of content selection, two problems often plague courses: First one was "The Potpourri Problem." Many elements from each discipline became a sampling of knowledge, which would result in random samplings of knowledge, lack of focus, and absence of structures of knowledge. The requirements of courses in an interdisciplinary program were carefully considered design features: a scope and sequence, a cognitive taxonomy to encourage thinking skills, behavioral indicators of attitudinal change. They must use both discipline-field-based and interdisciplinary experiences for students in the curriculum. Second was the "The Polarity Problem."

Interdisciplinarity and the discipline fields promoted a range of conflicts. Real tensions could emerge among instructors from various disciplines. Some instructors felt highly territorial about their own subjects and are threatened as new views of their subject were promoted. These two identified problems would affect the effectiveness of the program, we need to consider them when design the curriculum. To avoid these problems, it suggested that effective interdisciplinary programs: (1) had a continuous cognitive taxonomy to encourage thinking skills, behavioral indicators of attitudinal change; (2) used both discipline-field-based and interdisciplinary experiences for students in the curriculum(Jacobs, 1989). These two suggestions provided reference to set quality standards of interdisciplinary program. An interdisciplinary approach was often settled in a complex system, which made the quality standards very difficult. In Jacobs' s research(1989), four criteria were important for in the design process which could called intellectual criteria. The first one was validity within the disciplines which meant whether the identified concepts within each discipline that relate to a proposed interdisciplinary topic. The second one is validity for the disciplines, it meant comparing a concept from one subject with a similar one from another subject, the student might actually learn these concepts better than if they had been taught separately. The instructors and students achieved a better grasp of related concepts from different subjects by examining each one through multiples perspectives. The third one was validity beyond the disciplines, it indicated the case for or against curriculum integration has been made in relation to the lens of the distinct disciplines, the achievement of students might be enhanced if cross-curricular connections clearly improved vital subject-based ideas. The final criterion was

contribution to broader outcomes. Interdisciplinary education might help to shape the learner's overall approach to knowledge. Students may become more competent at and comfortable with absorbable thinking and with embracing multiple perspectives, for example. they might understand the theme of the program and may deeper comprehend the limitations in each discipline.

Regarding the booming of interdisciplinary programs, few researchers have systematically studied interdisciplinary curriculum (Lattuca et al., 2004). Like most university area doing now, it is obvious that we need to actively to build programs of research and specialized advanced teaching to foster experts. We also have to understand is that the general citizenry be broadly educated with respect to national goals and personal value choices is also important. From this perspectives, interdisciplinary program and general education share the same core. We could conclude that the discipline-based general education which is most accepted could be improved upon the introduction of multiple perspectives upon specific issues in order to exercise, skills of comparison, contrast, analysis, and above all, synthesis and integrate.

Marie-France Orillion’s article examines the relationship between interdisciplinary curriculum and student outcomes(2009). The greatest benefit students get from the interdisciplinary program is the ability to integrate knowledge though integrative learning. The word of “integration” in modern history happened in books theory of integrated instructions. Integration of learning was the “demonstrated ability to connect, apply, and/or synthesize information coherently from disparate contexts and perspectives, and make use of these new insights in multiple contexts(Barber, 2012). ”

Youatt and Wilcox (Youatt & Wilcox, 2008) took Michigan State University as an example, defined integrative learning as linking knowledge and competences from multiple sources in various and complex settings, or in brief, connecting knowing and action. Some other researchers claimed that “Integrative learning” is a broader of the two of integrative learning and interdisciplinary studies which was oversight term for constructions, approaches, and actions that bridge different divisions, such as general education and the major, introductory, and advanced levels, events inside and outside the classroom, theory and practice, and disciplines and other fields. Especially the general education programs, the extreme isolated disciplinary specialization had resulted in them that lack coherence and weakly prepare students for their roles in a global context whose main goal is to foster eligible citizens for the society.(Johnson et al., 2004; Stark, J. S., & Lattuca, 1997).

From the previous research, integration is one of the biggest takeaways when students participated in the interdisciplinary program, which is also a core benefit through the approach of interdisciplinary studies. Repko’s(2008) research was summarized into sections representing procedures in the interdisciplinary research process: (a) identifying relevant disciplines based on the , (b) developing competence in relevant disciplines, (c) analyzing the problem and evaluating each insight into it, (d) detecting conflicts in insights, (e) creating (or discovering) common ground, and (f) integrating insights and producing an interdisciplinary understanding.

Integration does not happen in vacuum, integrated course clusters were good context for student to connect, integrate and apply, knowledge learned from different

disciplines remind students to foster deeply think of the topics. The specialty in an interdisciplinary field creates obstacles for further integration. Some researcher was also worried that new interdisciplinary instructors "lack both interest and expertise in interdisciplinary research practice"(Jones, 2009). Integrated course clusters are an "expanded form of the linked course model" in which isolated courses from different fields were linked by "common themes, historical periods, issues or problems" (Smith et al., 2009) and were constructed together to form a "cluster" or "theme". For the integrated course clusters, several levels of integration were considered in the literature.

In clustered classes, the instructors choose a theme to attain considerable joint point between some of the topics in the classes, allowing students to experience a thorough cross-disciplinary foundation(Dinan, 2018). If the different courses a team has designed a "model" course when the requirement was adopted, the faculty actually teaching the course need a specific topic with which they feel comfortable. The process of tailoring the course to student's interests ensures that they get some exposure to other viewpoints on the theme and that their discipline's perspectives are represented. It is also benefit for integration to have all sections meet repeatedly for a common lecture; students get some exposure to each discipline's perspective from a supporter of the particular discipline. The lecture also provides a regular point of contact among sessions, giving students and faculty the sense that they are part of a larger, cohesive course. The images of the curriculum, in turn, reveal a new emphasis on integrating, connecting, linking, and clustering, interdisciplinarity became priority for more than one reason and addresses more than one kind of problem (Klein, 1999).

Integration of Knowledge

Integration is an exclusive character of interdisciplinary program. Thematic integration was the first level of the integration in a clustered theme interdisciplinary program (Burton, 2001). It meant selecting a theme or topic then selecting courses for knowledge and skills believed to be of help to students in understanding the theme, it was the key factor of validity within the disciplines. The representatives of various stakeholders for example instructors should gather to determine what components of their disciplines could clear up the themes under study based on the based topics. In this level of integration, themes or topics became the priority of curriculum organizers, course construction, and knowledge and skill centered on it was used to drive the development of a curriculum design. In an article early in 90s (Lipson et al., 1993), explained by themes could provide a framework of coherence for both teachers and students, the continuousness of knowledge across different fields would keep students from thinking. Another additional big advantage of organizing thematical curriculum could benefit for promoting metacognitive awareness (Pappas et al., 1990). Because giving more opportunities for students to have reflective thinking that provided by additional time and the breadth and depth of study, the process permitted students to become unambiguously informed of the knowledge and skills needed to learn (Lipson et al., 1993). Thematic teaching and learning could promote the attainment of an integrated knowledge base. After the thematic course's construction was established in that setting foundation of the base, integration comes to the knowledge itself. Take the matter of renting eviction, it might make sense to analyze the elements in the lives of individuals and in society at

large that relate to income and its causes. This process needs involves differentiating, integrating, reprioritize, and restructuring ideas.

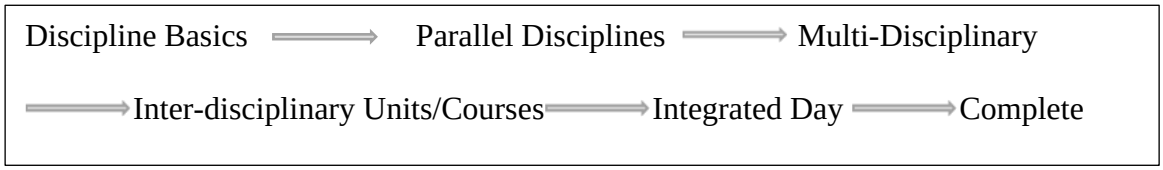
"Knowledge integration is the dynamic process of linking, connecting, distinguishing, organizing and, structuring 'models' of phenomena" (Linn, 2000). In 1995, Marcia C. Linn (and collaborators) proposed scaffolded knowledge integration (SKI) framework which is an instructional design model to enhance science teaching in school classrooms, now was used in many science education settings. The framework has four main elements: (a) identifying new goals for learning, involves introducing a bunch of models to build student insights and encourage testing, revising, and reformulating scientific ideas. For knowledge integration, students need to understand the process of thinking independently. (b) making thinking visible, this element emphasizes making alternative models accessible to students. Through this process, students need to understand several perspectives on scientific ideas in order to experience the procedure of comparing and testing scientific ideas. (c) encouraging lifelong learning, and (d) providing social supports, this element involves coordinating productive social interactions in the classroom can facilitate each other compare ideas and connect models when they respect each other's difference (Linn, 1995). The whole framework combined network resources and software with sound pedagogical principles to improve science learning (Bell et al., 1995). Although SKI model was mainly for science education, the whole framework includes techniques for helping students to build the process of students' own reflection of ideas and knowledge, interaction with peers, instructions which were also could be utilized in the process of knowledge integration in other fields.

It also provided a firm foundation to show the knowledge integration could improve learning and are being deployed more widely in varied school settings. Furthermore, as global citizens, the skills learned in knowledge integration will help students develop skills they will need throughout their lives, and also provided tools for the development of integrated understanding of complex world now.

SKI initiated a very clear process of integration in science education, and it also made it pointed out that "to carry out interdisciplinary research, one must have both disciplinary capability and interdisciplinary conversance" (Hartesveldt & Giordan, 2008). Student work in the interdisciplinary program was to integrate the disciplinary concepts and to provide a window into a learner's ability to pull from multiple areas of learning in analyzing and integrating knowledge in different disciplines. These elements of learning in any fields were considered the archetype of learning and acquirement of concepts relevant to the discipline. Jacobs (1989) also proposed a continuum for the interdisciplinary curriculum design options:

Figure 1

Continuum for the Interdisciplinary Curriculum Design Options



Different approaches had its own different advantages and disadvantages, but we had its concluded in all the approaches, the disciplined foundation was important in the process of interdisciplinary curriculum. "Disciplined field was a specific body of teachable knowledge with its own background of education, training, procedures,

methods, and content areas" (Piaget 1972). There were distinctive frames of character reference and types of statements, and each of these suggests unique procedures and end results in the particular discipline fields. Arthur R. King and John A. Brownell(1966) in The Curriculum and the Disciplines of Knowledge: A Theory of Curriculum Practice offered ten characteristics of a discipline of knowledge, these characteristics were a prerequisite for interdisciplinary curriculum design. It helped to conclude that at the first step, disciplined knowledge helped to find common grounds for discussion.

Integrating disciplinary insights into a comprehensive understanding needed many steps including identifying and evaluating, resolving and constructing, creating, producing and testing (Newell, 2001). In this development, disciplines may conflict each other by their own foundation. To resolve the conflicts may happen between disciplines, the structures of the disciplines had to be known and understood before true connections and ties can be processed. From a work inductively conducted from transcripts of interdisciplinary classroom conversations and directed by a Bakhtinian view of language and cognition concluded the connections were made (or not made) by individuals who identified similarities and differences among various disciplinary contents, propositions, classroom genres, and ways of knowing (Nowacek, 2005), which meant discipline grounding is the basic of integration. In Boix Mansilla and Dawes Duraisingh ' s (2007) research, this criteria was called as "disciplinary grounding", it considered the appropriateness of disciplines and application consistent with "disciplinary theories, findings, examples, methods, validation criteria, genres, and forms of communication" (p. 222) which means whether any main disciplinary perspectives were ignored and

confirmed that the instruction process didn't exhibit any misunderstandings. Relationships among content fields, processes, and products need to be started to investigate in the context of the disciplines from which they have been drawn (Burton, 2001). A well-established disciplinary grounding diminishes the risk of the linked courses being "a mile wide and an inch deep" or "a light educational experience"(Mansilla & Duraising, 2007). When different fields of knowledge and skill were unique to a particular discipline, such areas must be presented even if no real linkages could be discovered. As a result, structures of learning within the disciplines that need to be presented at appropriate times take priority over integrated approaches which share the same meaning with discipline grounding in the previous literature. A major problem with interdisciplinary curriculum programs should be avoided was the instructor attempted to present knowledge and skills unique to disciplines in which they have little expertise(L. H. Burton, 2001). This perspective was endorsed by Holley (2017), who indicated that interdisciplinarity did not reduce the role of the discipline in education(p. 1) but rather recognizes the lack of boundaries in knowledge creation, which understands that knowledge to extend beyond predetermined or normative silos.

After consideration of the disciplinary grounding, it came to the integration process, which was the essence of interdisciplinarity program(Borrego & Newswander, 2010). When interactive and connective relationships (also referred to as joint points and linkages) are established between the knowledge and skills in two or more, the knowledge integration would happen. In most discussions of interdisciplinary learning outcomes. there were still many different opinions. Knowledge integration was viewed as

a higher-level integration than thematic integration. The Association of American Colleges and Universities (AAC&U) indicated "significant knowledge within individual disciplines serve as the foundation, but integrative learning goes beyond academic boundaries" (2004, p. 1). The integrative learning came in a main way: connecting skills from multiple resources. Julie Klein and Newell (1996) defined interdisciplinary study as "a process of answering a question, solving a problem, or addressing a topic that is too broad or complex to be dealt with adequately by a single discipline or profession. IDS draws on disciplinary perspectives and integrates their insights through construction of a more comprehensive perspective" (pp. 393-394). The main point was students construct the joint point of knowledge such as multiple viewpoints. Some research (Songer & Linn, 1991a) showed the development of scientific knowledge involved student integrating ideas about separate events, and recognizing that outwardly not correlated events were in fact could be described by the same set of concepts. This existed research showed when learners were having a solid disciplinary grounding and realized the connections between knowledge, then the integration might happen.

Integration across Disciplines

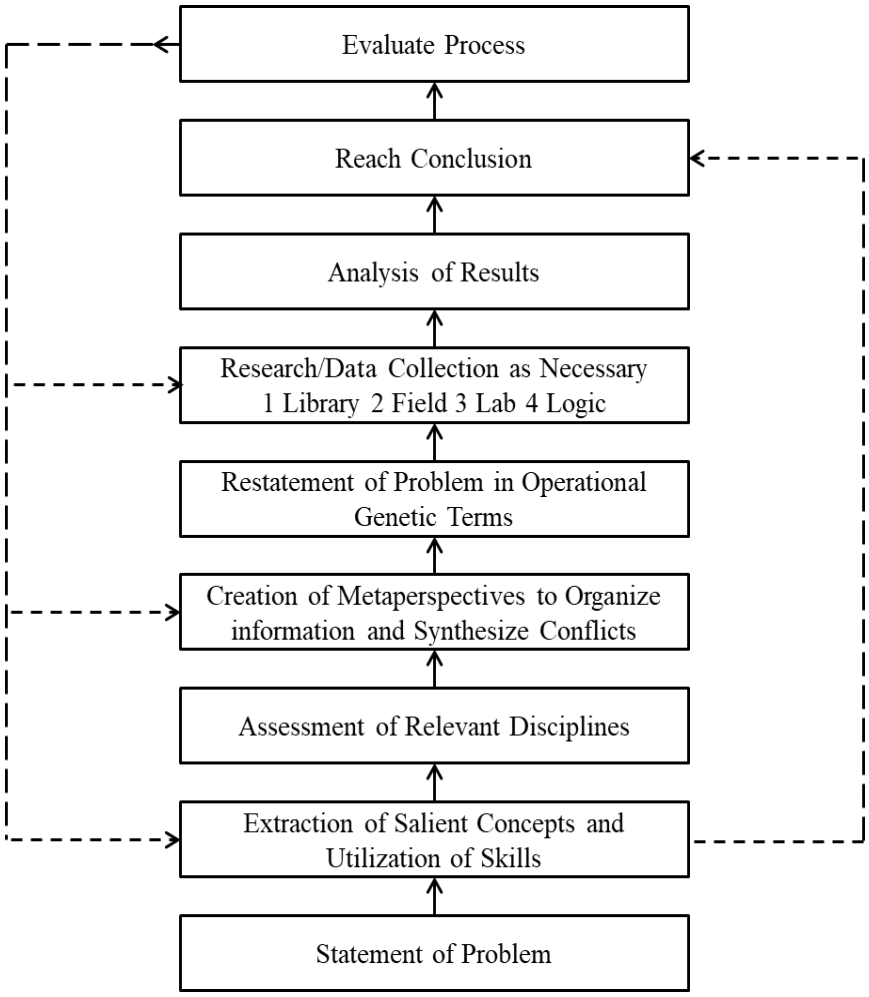
Integrate knowledge may happen next across the discipline boarders. Tondreau (2003) suggested that development arises from the ways in which individuals made sense of changes and dissonant information, and constructed a meaningful narrative. It meant the second level centers on a more thorough integration of the different perspectives identified by the different disciplines involved. This means, by blending inputs from

more than one discipline, students can reflect conclusions and finally move toward a more inclusive understanding of the problem at hand (Hursh et al., 1983).

There are many ways to make connections across disciplines. Nowacek (2005) work concluded several approaches to integrate, vocabulary introduced in one discipline while considering another: apply materials from one discipline as an example, making comparisons between materials from two different courses; using material from one course to reevaluate material from another. Thematic integration belongs to the level of the construction of the curriculum. The integration beyond thematic integration involved learners' participation. In 2010, AAC&U stated, "Integrative learning was a method to connect relevant experiences and academic knowledge; seeing and making connections across disciplines and perspectives; adapting and applying skills, abilities, theories, or methodologies gained in one situation to new situations; communicating in language that demonstrates cross-disciplinary fluency; and demonstrating a developing sense of self as a learner, building on prior experiences to respond to new and challenging contexts".

Figure 2

Process for Interdisciplinary Study of a Given Problem



Note. From An Interdisciplinary Model to Implement General Education by Barbara Hursh, Paul Haas and Michael Moore, 1983. *The Journal of Higher Education* (Vol. 54, No. 1), Published by: Taylor & Francis, Ltd. CopyRight 1983 by Taylor & Francis.

This definition widens the integration context from class setting to student's active engagement traditional research on classroom experiences fail to consider

students' experiences outside class nor their personal background characteristics which was called. This highest level of integration was called Learner-initiated integration, was believed to be the most important in this explosive knowledge era because it relates to the employment of higher-level thinking skills throughout life. It happened when learners realized connections from separate fielded instruction. Based on this, the curriculum of programs should be designed to guide learners to independently integrate with information and lead students to discover connections. Instructors should learn to ask questions to provoke students to find connections, value independent and critical thinking develop consecutive understandings in separate disciplines of knowledge and skills; and to establish thought patterns that lead them to look for connective relationships across all disciplines of learning (Burton, 2001).

Through the leaner-initiated integration, students became independent, confident learners who 'learn how to learn' and develop lifelong learning skills, broad knowledge brought impact to student's futures in the way that "Their cognitive development allows them to see relationships among content areas and understand principles that cross curricular lines. Their psychosocial development gives them the ability to understand people and to look at situations from various viewpoints" (Duerr, 2008, p.177). The learner-initiated integration was also an integration of methodology and pedagogy, and a much-needed lifelong learning skill. The skills that interdisciplinary courses provided were so valuable to students' future that sought out by colleges and businesses. Learners also benefited from solving real-world related problems, offering multiple solutions (Association of American Colleges and Universities & The Carnegie Foundation for the

Advancement of Teaching, 2004). The integration of interdisciplinary could lead to a future of discovery and innovation (Youngblood, 2007). At the same time, the integration enhanced both the teaching and learning of the disciplines in some context throughout a lifetime, because Students and instructors involved in advancing critical thinking, communication, creativity, pedagogy, and essential academia in particular interdisciplinary programs (Jones, 2009; Taylor, 2008).

The AAC&U definition of integration also mentioned "developing sense of self as a learner, building on prior experiences to respond to new and challenging contexts". The concept that learners brought their own prior knowledge and experiences to learning environments was the base stone of educational concept of "development from within" intrigued by John Dewey's (Dewey, 1933, 1938), it concluded interaction between the learner's prior knowledge and experience and what is being learned make meaning in the process of education. Dewey and others (Bowers, 1967; Zilversmit, 1993) proposed progressive education that promotes integrated understanding through open investigation, progressive learning experiences emphasizing privilege experience over learning, interaction over silence, applied learning over isolated research and instruction, and *courses that integrate rather than isolate academic disciplines to make learning more meaningful*. The natural act of telling one's own story as essential in making the link between lived experiences and academic content for students (Rossiter & Clark, 2007), in the process of making links, higher level integration ability was processing.

Integrative Learning

While *integration* is a primary learning goal, *integrative learning* is a broader term with origins in contemporary learning sciences moving students from memorizing facts to more complex learning that is inclusive of constructing knowledge, setting learning strategies, examining process, and experiencing a range of activities to achieve the learning goal of integration in an interdisciplinary program (Klein, 2005). Conceptually, integrative learning constructs a bridge to span gaps across the college experience including general education and the major, introductory and advanced levels, experiences inside and outside the classroom, theory and practice, and disciplines and fields. As previously mentioned, what happens inside a classroom is often considered distinct from what occurs outside a classroom, and learning is only synonymous with the classroom experience. The dynamic nature of higher education today, which has responded to pressures to address complex social and scientific issues, results in institutions paying more attention to non-academic and academic experiences to promote students' learning in both academic and career domains. Integrative learning in college is a combination of students' experiences both inside and outside the classroom which help the advancement of desired skills of undergraduates (Awang-Hashim et al., 2021). For example, the AAC&U proposed learning outcomes around "work, life, and citizenship" to arm students with essential skills and proficiencies to solve problems, revolutionize innovative solutions and integrate with diversity in community (National Leadership Council For Liberal Education & America's Promise, 2007, p.2). The findings suggested that the relationship between students' learning outcomes and interdisciplinary

curriculum was reconciled by environment and culture at university. To understand the relationship between student outcomes and the curriculum, it was essential to explore the environment as it is enacted in classrooms (Marie-France Orillion, 2009). The learning environment affected students learning experiences and outcomes such as communication skills which should also be considered organizing an integrated course cluster learning community that includes a common theme across courses(Barnett et al., 2009).

Learning environments provoked higher levels of learning by offering the discussion of views, allowing for the united construction of knowledge, and sharing of the cognitive capacity (Dillenbourg, 1999; Ploetzner et al., 1999), different variables in learning environment that affect students' course satisfaction and improvements in learning outcomes, Songer and Linn (Songer & Linn, 1991b) also found that students believes had impact on the process of knowledge integration, thereby encouraged students to engage with a deeper level thoughts (Anderson, 2003).

A supportive environment interaction with others in the learning environment has been shown to contribute to the development of students' knowledge, thoughts, and values (Hartup, 1989; K. Smith, 1979). In the vast majority of cases, students were integrating on their own or with their peers, faculty members, student affairs professional, or other adult mentors. It was widely accepted that faculty-student interaction was a positive predictor of students' academic achievement. Astin indicated in In What Matters in College (1993) "student-faculty interaction has significant positive correlations with every academic attainment outcome" (p. 383). From Dewey's reflective thinking to Bruner's discovery method, "teaching techniques that involve experiential, collaborative,

and reflective learning that relate the course content to real-world issues" was vital to student engagement (Roszkowski, 2003), which can make student cognitively "engaged", it meant the student willingly commenced while learning in the context of an learning task (Chi & Wylie, 2014; Corno & Mandinach, 1983). Callaghan and Bower (Callaghan & Bower, 2012) emphasized the key role of the instructors in designing engaging and effective online learning in social networking environments. They argued that the environments stimulated student motivation and engagement and enable students to use higher order thinking skills in the learning. Active, collaborative learning and faculty-student interaction promote academic accomplishment and student retention rate. In a very early time, Snow (1973) demonstrated if an instructor had an interactionist conception of the relationship with students, he would take an active role to build an interactive relationship, both student and the instructor would get cognitive gain from the relationship. Instructors could also promote the growth of thinking skills by incorporating active learning strategies into the classroom, collaborative learning were a good strategy to promote student-faculty interaction which was benefit for students thinking skills (Kim et al., 2013; Bjorklund et al., 2004). In addition to this, faculty-student interaction and feedback was great contributor to the other learning outcomes such as group skills, occupational awareness and engineer competence(Bjorklund et al., 2004). The National Survey of Student Engagement (NSEE), which is the one of the biggest student experiences survey across the county, includes a section called "Student-Faculty Interaction." The NSSE measures institutional achievement at generating this kind of interaction and compares the results between universities and the peer institutions.

Faculty-student interaction occurred on many forms and many degrees of frequency, it could happened both in and out of the classroom(Bjorklund et al., 2004). The interactive components ranging from one-on-one interaction to group-based discussion activities such as discuss course-related topics, offer academic advice, or informal discussion about students' feeling about the learning. Both faculty and students could sense the gain of out-of-classroom engagement as revealed in interview data collected at SUNY Oswego (Bosch et al., 2008, p.89). Other research focused on instructors' functions in student learning process, and presents some suggestions for advancing the ways in which faculty interact with and offer response to students (Bjorklund et al., 2004), this kind of reflective learning which highlighted the faculty-student interactions including reflective assignments are given, instructors respond actively to students experiencing questions or problems (Roszkowski, 2003)."Knowledge and the process of learning as distributed among learners, instructors, and learning resources located within the context of an activity" (Polin, 2004).

To make students more engaged into their learning and get a better satisfaction rate, peer interaction was also a vital factor. College outcomes researchers had amassed empirical evidence of peer influence on learning (Pascarella & Terenzini, 2005), LaPointe and Gunawardena (2004) showed that peer interaction on self-reported learning outcomes implied that peer interaction has a powerful and direct effect on student learning outcomes. Students' characteristics and behaviors influenced other students' behavior with conventionally measured academic test (like SAT) influencing conventionally measured academic performance (e.g. GPA), operated equally and

symmetrically across characteristics and behaviors, resulting stratification(Hoxby et al., 2013). Student cooperative learning such as peer-group also increased the student involvement (Collier, 1980). All group members were supposed to achieve a shared set of goals, each member was expected to make their own part of contribution to the work by sharing their opinions, helping to solve problems, arguing rationally to reach a final agreement. This working way was good to change the student's role from passive to active. Through this process, the intellectual stimulation within the process is extremely high (Collier, 1980). Not only large group, but also small semi-independent group also positively stroke students learning outcomes of higher order skill (Collier, 1980). There were many other formats of peer interaction. Peer teaching and peer tutoring programs had been integrated with different programs to help provide better assessment in the classroom, and improve learning outcomes of students and afford valuable information to other stakeholders (Miller, 2000; Topping & Ehly, 1998). Peer assessment, as an important approach used in peer teaching, was one of peer collaborative learning approaches, requiring students simply to assess one another's work by relevant criteria, and to provide feedback(van den Berg et al., 2006). In the process of peer teaching or peer tutoring, the important point it the interactions that are formed between or among tutors, students, and instructors. More closely examining the socialization with a focus on everyday interactions is essential (Colvin, 2007).

The previous research concluded social engagement and integration outside the classrooms, at some extent, affected student achievement(Pascarella & Terenzini, 2005; Wolf-Wendel et al., 2009). From a broader perspective, peer-interaction was closely

correlated to the goals of general education not only on the academic dimension but also social responsibilities. Emerging student on the peer interaction with diverse background was part of a long-term effort of general education, which was to prepare the next generation of citizens for a multicultural society today. It could bring the leadership with a stronger social awareness and the high-order thinking abilities to alleviate social problems related under the complexities of inequality (Hurtado, 2007). Students were different from each other, as learners, they chose their strategies based on their own characteristics. Some students consciously pursued the intellectual, social, and work-related activities that require increasing levels of critical thinking, leadership, and problem solving, but some were not. Establishing a peer culture formed the mesosystem which included roommates friendship group, job and classmates would provide contextualized explorations of interactions among and between students and environments (Renn & Arnold, 2003). Under today's global discourses, students with multiple cultural backgrounds, would bring diverse knowledge perspectives and the variety of meanings to the classes.

With the widely acceptance of technology in class, social media offered opportunities for students to interact and share instant user-generated content, it changed the traditional format of peer-interactions. Student could learn knowledge whenever and wherever they like, and of course, they could get in touch with instructors and other students whenever they want. This flexibility made conversations and learning occur in varying ways. Learning experiences involving social media had helped to inform students on their own learning design(Casey & Wells, 2015).

The interdisciplinary courses was also a team-taught enhancement of student performance (Youngblood, 2007), Repko (2008) explained in more detailed:

> Interdisciplinarity is often a collaborative process……An expert interdisciplinarian is one who is able to integrate the input of others to address an issue, which may include coordinating team members. This trait applies especially to interdisciplinarians engaged in technical and scientific studies that most commonly involve teamwork. (p. 44)

To promote student more engaged in the program, this study will take their academic and social interaction into consideration.

Designing Model to Measure Interdisciplinary Competence

The previous review concluded the historical development of general education, the characteristics of the interdisciplinary program, the assessment of learning outcomes. This part will address frameworks that are particularly established for an interdisciplinary program.

The Interdisciplinary Concept Model (Jacobs & Borland, 1986) suggested a framework for curriculum development with interdisciplinary insights at its primary, the stakeholders could brainstorm and assess subjects and disciplines that might be contained in an interdisciplinary program. Boix Mansilla (2010) summarized four cognitive procedure that which may improve student learning outcomes for an interdisciplinary program: 1) establishing the program purpose; 2) balancing disciplinary grounding; 3) establishing integrations; and 4) keeping a critical attitude. In order to stimulate these whole processes, students' work should be based on both disciplinary foundation and

interdisciplinary insights. From these we can clearly see it may take a significant amount of time for students to develop sufficient levels of complexity in these interdisciplinary cognitive processes. It is essential to ensure that the assessment of learning outcomes from interdisciplinary describes the complexity of the cognitive and metacognitive work including a transdisciplinary course or program (Ashby & Exter, 2019). Boix Mansilla and Duraisingh (2007) developed a framework aimed at establishing a culture of interdisciplinary program evaluation. This framework includes important criteria to measure the learning outcomes from interdisciplinarity: (1) Solid foundation or grounding in disciplines; (2) Progression through integrating multiple disciplinary perspectives; (3) Critical thinking of the way to synthesize disciplinary knowledge. Aligned with Boxi Mansilla and Duraiginh's framework in 2007, Lattuca et al. (2013) expanded a set of criteria to measure interdisciplinary competency for students in engineering area. It outlines eight key factors such as awareness of both disciplinarity and non-disciplinary perspectives, then finding the common ground through knowing of disciplinary limitations, understanding the biases through reflexivity, finally gain the integrative competency. The Bigg's Model of Constructive Alignment (2003) is an learning outcome-based model and has been endorsed as an operational framework for developing assessment of interdisciplinary learning experiences (Stefani, 2008; Yang, 2009). It emphasizes three factors: intended learning outcomes which refers as the learning environment, teaching and learning activities including the experiences in the program, assessment tasks.

Literature Gap

The literature review firstly brushed the history of general education, the chapter traced it back to the original foundation of general education, tracked the development of general education in the historical context, to see the process of formation of general education today. From this process, we clearly understood general education derived from liberal arts education, but the goal of general education programs today is more complex than the past. It has some overlap with liberal arts education for example to foster citizens and broad students' width of knowledge, but it will also combine more professional knowledge, provide a wide foundation for students' professional career. The existing research on general education informed us the core purpose the programs today, also helped us to define general education in a new global context. General education needs to change its format to adjust to the higher education today.

Then the chapter reviewed the existing research and instrument designed to measure student learning outcomes and evaluate general education program. Outcome's assessment can tell the stakeholders whether the project has achieved its goals, whereas a process assessment can inform about why such results occurred. Since students get different learning outcomes according to the levels and programs, different instruments provided us various inspiration for the future instrument development from the approaches to item content. Nowadays, most student learning outcomes assessment in US was no longer driven primarily by governmental accountability, but by institutional level use for guiding decision-making, curricular change, and improving teaching or student learning (Kuh et al., 2014). The report revealed the essentiality of internal accountability,

to develop an instrument to assess student learning outcomes from institutional or program themselves will provide decision-making support for different stakeholders.

The literature also dug into the development of interdisciplinary program, explained the advantages and disadvantages of interdisciplinarity from philosophic and practical perspectives, and concluded the key indicators in interdisciplinary program that affect integrative learning in a program. The process of integrative learning was also a focus of literature which revealed the key factors that will affect students' learning outcomes.

However, a literature review is absolutely limited to the understandings of those who have their own experiences and knowledge, further steps can still to be taken in this research. The (table 2-1) shows the mainstream instruments to assess student learning outcomes for the general education, we can conclude that the mainstream instruments which to assess general education were primarily assess detailed ability that the general education emphasized in the past. In other words, to measure students' growth in skills such as critical thinking or writing without an integrative insight. But from the program level, there were few instruments considering the curriculum construction, the knowledge continuousness, the learning environment provided the program. The learning outcomes students get from a program is not an isolated skill but an integrative learning result. This research fills the gap of literature by considering integrative learning as a core conception, to assess the learning outcomes as a continuous result from a program level.

Second, the existing literature did not explore enough the relationship between general education and interdisciplinary program. General education was a distribution

requirement across the arts, humanities, math, sciences, and social sciences (J. Fredericks Volkwein, 2009), it perfectly fits in interdisciplinary approach. Although interdisciplinary program and general education share natural common goals based on their characteristics, most of the time, general education courses at institutions are considered as isolated courses just like small islands in a big see, but not as a whole interdisciplinary program, or clustered thematic program. As a result, students will not get continuous knowledge in depth, when assess the learning outcomes of general education, administrator will not take the continuousness of course construction into consideration. This research will put more attention to the current status of assessment for general education, emphasize the importance of the inner connection between the courses that the existing literature didn't focus on.

Third, the clustered thematic classes are very special types of interdisciplinary program. A common theme makes the courses in general education become more connected with each other; it deepens students understanding of a social issue from different perspectives. Student are integrating learning across disciplinary settings. This learning experiences satisfy the educational requirements of general education nowadays, provide the most important takeaways for professional development such as how to integrate academic knowledge to real world. This is also the most important gap this research is going to fill, developing an instrument to assess the integrative learning in the learning process.

Finally, for most traditional assessment methods, in many cases, instructors or students might not access the feedback on their work, and adjust the feedback and

actively use it to inform future strategies to their learning (Dilly Fung, 2017). In this research, all the practice results had been provided to the leader, and students opinions on minor will be considered as the evidence to revise the instrument.

Chapter 3 Methodology

The purpose of this study is developing an instrument to measure the learning outcomes of integration in a progressive general education program. To achieve this purpose the following research questions are used to guide this study. This chapter will cover the methodology used to address these questions. It will include the establishment of the instrument, the instrumentation process, two waves of data collection process (pilot study and filed study data collection) and concludes with a data analysis plan for the two waves of data including descriptive analysis, initial item analysis, and qualitative data analysis.

Instrumentation

Approaches to Testing

Classical Test Theory (CTT) and Item Response Theory (IRT) are the main approaches to interpreting test outputs. The two approaches have their own pre-assumptions and method of estimating parameters. CTT and IRT model data as a continuous variable while latent variable analysis models data as categorical variables. Second, CTT and IRT in most cases believed that all items measure the same latent variable, the individuals' responses are applied to assess their capability align with some continuous scale. The final difference is that in CTT, total test scores are used to make inferences about the structure in an instrument, IRT analyses responses at the item level.

Overview of CTT and IRT

Classical Test Theory. Classical test theory (CTT) is a theoretical framework used to evaluate a measurement instrument. It was introduced in the early 20 century by

Spearman (1907, 1913) to distinguish why test items or scores are imperfect to measures the goal (Crocker & Algina, 1986). By examining the relation between true score(T), observed score(X) and error score (E) (Spearman, 1907, 1913). This linear equation is expressed as:

$$X = T + E \quad (1)$$

the observed score is one's actual test score and reflects the linear combination of true score and error score (Spearman, 1907, 1913), true score T is recognized as the expected test score over parallel testing. The fundamental CTT has three assumptions which are required in the process of evaluating the measurement objects. First, is uncorrelated true score and error scores for the whole population. Second, the expected error scores value in the population should be assessed as zero over continued and repeated testing. Third, the error scores on the parallel tests are uncorrelated (Nunnaly & Bernstein, 1994; Traub, 1997). Although CTT is strict with assumptions, one reason to continue to use CTT as one method is because CTT can be adopted when the testing has a comparatively smaller sample size.

Three types of item parameters are calculated in the CTT model: "parameters that describe the distribution of responses to a single item (i.e., the mean and variance of the item response), parameters that explain the degree of relationship between the item and some criterion, and indices that are a function of both item variance and relationship to a criterion" (Crocker & Algina, 1986, p.311). In a CTT framework, most frequently utilized indices are describe single item indices, with item difficulty and item discrimination (Ellis & Mead, 2002). Item difficulty indicator shows the proportion of

participants answering an item correctly. In the case of this research, it refers to endorse at or above a predefined response category (i.e., “6 =strongly agree”) is similar to answering an item correctly because the higher the level of integration in minor is, the better.

Item Response Theory. IRT analysis is a set of numerical models for measurement, it is also recognized as “latent trait theory” or “item characteristic curve theory”. It produces different indicators of describing and comparing items and measurement instruments, the result can lead items amendment or deletion by estimating parameters whether they are consistent with the goals of the test(Hambleton & Jones, 1993), it is an approach that depicts the correlation between an individual participant’s response to an item and the underlying trait (Embretson & Reise, 2013). The most often-used indicators, for example, Item discrimination, Item Characteristic Curve (ICC), the item response function (IRF). Item discrimination is assessing the relationship between the item and the total test score (Nunnaly & Bernstein, 1994), it identifies items in which high-scoring participants have a high possibility to correctly answer or endorsing an answer while low-scoring examinees have a opposite situation(Crocker & Algina, 1986). Item characteristic curves, also called as item characteristic function is a graphic technique which is used to picture the relationship of the response score that expected and the trait being measured, it presents the connection between the latent variable or trait being measured and the probability of responding an item correctly (Engelhard Jr, 2005).

$$f_i(\theta) \equiv p_i(\theta)^{ui} Q_i(\theta)^{1-ui} \qquad (2)$$

In this equation, *p* represents the probability of participants to respond the item correctly, when *Q = 1 – p*. *u* represents the dichotomous items 1 is the corrected response and 0 is the incorrect ones. In conclusion, the ICC represents the expected probability for a participant to select the correct response.

IRF is the indicator that sum up of ICCs for an item and gives information about the where and contribution information that each item along the ability continuum, it represents the correlation between a tester's ability and the probability to choose an item. The function can be written as equation 2 (Yen et al., 2006):

$$p_i(\theta) \equiv p_i(X_i = x_i|\{\theta\},\{\delta_i\}) \quad (3)$$

The most popular used models are one-parameter logistic (1PL), two-parameter logistic (2PL), and three-parameter logistic (3PL) models(Wang & Osterlind, 2013), it estimate one, two, or three of these parameters,

One-parameter logistic (1PL):

$$P(X_{is} = 1|\theta_s,\beta_i) = \frac{\exp(\theta_s-\beta_i)}{1+\exp(\theta_s-\beta_i)}, i = 1,2,\cdots,n \quad (4)$$

Two-parameter logistic (2-PL):

$$P(X_{is} = 1|\theta_s,\beta_i,\alpha_i) = \frac{\exp[\alpha_i(\theta_s-\beta_i)]}{1+\exp[\alpha_i(\theta_s-\beta_i)]}, i = 1,2,\cdots,n \quad (5)$$

Where, X_{is} = response of person s to item i (with response options 0 or 1)

θ_s = Latent trait level for person s

β_i = difficulty level for item i

α_i = discrimination power for item i

The 1- and 2-PL models show the mathematical probability of a correct response on item i to person ability (θ), item difficulty (β), and item discrimination (α).

Three-parameter logistic (3-PL)

$$P(X_{is} = 1|\theta_s, \beta_i, \alpha_i, c_i) = c_i + (1 - c_i)\frac{\exp[\alpha_i(\theta_s - \beta_i)]}{1 + \exp[\alpha_i(\theta_s - \beta_i)]}, i = 1,2,\cdots,n \quad (6)$$

where, X_{is} = response of person s to item i (with response options 0 or 1)

θ_s = Latent trait level for person s

β_i = difficulty level for item i

α_i = discrimination power for item i

c_i = random guessing factor for item i

Model fit indicates which IRT model predicts responses to a particular item and is calculated with either graphical or statistical approaches. The model fit considers for both individual item and the overall model. From individual item perspective, large standard errors usually imply problems. The 1-PL and 2-PL models are nested models, that is why they can be compared using the log-likelihood function, the Akaike Information Criterion (AIC) (Akaike, 1974) and the Bayesian Information Criterion (BIC) (Schwarz, 1978). AIC and BIC both describe this by changing a factor to see the number of parameters. The BIC loads the number of parameters by the natural log of the sample size. The equation of AIC and BIC are:

$$AIC = -2\ln(LL) + 2(kk) \quad (7)$$

$$BIC = -2\ln(LL) + \ln(n) * k \quad (8)$$

Where k is the number of parameters estimated by the model and n is the sample size. A smaller value for either indicator the better model fit.

Polytomous IRT Models. The items in this instrument are all Likert-type, the traditional IRT doesn't fit these items. Polytomous IRT models have been developed to estimate the probability of items which have more than two possible scores choice. There are two types of polytomous models: the first kind is divide-by-total models that the probability of response in either category k or category k-1. Second are the models the probability of response in category k or above. If item i is scored 0, 1…, mi. The item has Ki response categories, where Ki= mi +1. For x=j, the probability of response x=j for item will described below.

Partial Credit Model or PCM (Masters, 1982, 2016) is a model using items which have more than two ordered response categories. The probabilities of responses are dependent on each other within an item and sum to 1, which means the respondent will always have a response to the item, even if it is zero. The PCM models have many application ways. Normal items are either graded on a scale regarding either the degree of agreement with the statement about the performance of an individual, or which the rater identifies the rated individual as showing a skill or learning outcomes. In most cases, a lower score therefore suggests a lower degree of the competence. Similarly, higher observed scores exhibiting the behavior more frequently. Utilizing the PCM to these types of assessment instrument rating for example student experiences assumes that obtaining a high score reflects students more satisfied with their experiences than

receiving a low score. If item i is scored 0, 1,…,mi . The item has Ki response categories, where Ki= mi +1. For x=j, the probability of response x=j for item i is:

$$P_{six}(\theta) = \frac{exp\left[\sum_{j=0}^{x}(\theta_s - (\beta_i - \tau_{ij}))\right]}{\sum_{r=0}^{m_j}\left[exp\left[\sum_{j=0}^{r}(\theta_s - (\beta_i - \tau_{ij}))\right]\right]} \quad (9)$$

Where $\sum_{j=0}^{0}(\theta - \delta_{ij}) \equiv 0$

The Rating Scale Model, or we call it RSM(Andrich, 1978), it is written as the following equation. The RSM models can be applied to the Likert items instrument which elicit the same rating structure across items. If different rating scales used for each item or the same number of rating categories that carry different meaning across different items, the item formats may suggest the use of partial credit scoring models.

$$P_{six}(\beta) = \frac{\exp\left[\sum_{j=0}^{x}(\theta_s - (\beta_i - \tau_j))\right]}{\sum_{r=0}^{m_j}\left[\exp\sum_{j=0}^{r}(\theta_s - (\beta_i - \tau_j))\right]} \quad (10)$$

Where $\sum_{j=0}^{0}(\theta_s - (\beta_i - \tau_j)) \equiv 0$

The “x” in the equations the score and “s” is for person in this formula.

The Generalized Partial Credit model (GPCM) was devised by Muraki in 1992 (Muraki, 1992) based on Masters' partial credit model (PCM) by adjusting the assumption of uniform discrimination of items(Muraki, 1997). It assumes that the (Bock, 1972)dichotomous response model.

$$P_{six}(\beta) = \frac{\exp\left[\sum_{j=0}^{x}\alpha_i(\beta_s - (\delta_i + \tau_j))\right]}{\sum_{r=0}^{m_j}\left[\exp\sum_{j=0}^{r}\alpha_i(\beta_s - (\delta_i + \tau_j))\right]} \quad (11)$$

Where $\sum_{j=0}^{0}\alpha_i(\beta_s - \delta_i - \tau_j) \equiv 0$

Nominal Response Model (Bock, 1972) includes a big serious of functions appropriate for statistical description of individual qualitative behaviors, it was formulated for specifying the probability of in one of several mutually exclusive and exhaustive categories(Bock, 1997).

$$P_{ix}(\theta) = \frac{\exp\left[\sum_{j=0}^{x} \alpha_i(\theta_s - \delta_j)\right]}{\sum_{r=0}^{m_j}\left[\exp \sum_{j=0}^{r} \alpha_i(\theta_s - \delta_{ij})\right]} \tag{12}$$

$$= \frac{\exp\left[\sum_{j=0}^{x} \alpha_i(\theta_s - (\beta_i - \tau_j))\right]}{\sum_{r=0}^{m_j}\left[\exp \sum_{j=0}^{r} \alpha_i(\theta_s - (\beta_i - \tau_j))\right]}$$

$$\sum_{j=0}^{0} \alpha_i(\theta - \beta_i + \tau_j) \equiv 0 \tag{13}$$

Graded Response Model is a model that each response is generally modeled as a 2PL model, because it considers each ordered option as a dichotomous item (Samejima, 1969). The graded response model includes a group of statistical models that deals with ordered polytomous categories, the categories include rating such as Likert scale strongly disagree, disagree, agree, and strongly agree, commonly used in surveys.

$$P \geq 0 = P_0 = 1.0 \tag{14}$$

$$P(x \geq 1) = P_1 = \frac{e^{\alpha_i(\theta - \beta_{i1})}}{1 + e^{\alpha_i(\theta - \beta_{i1})}} \tag{15}$$

$$P(x \geq 2) = P_2 = \frac{e^{\alpha_i(\theta - \beta_{i2})}}{1 + e^{\alpha_i(\theta - \beta_{i2})}} \tag{16}$$

$$P(x \geq 3) = P_3 = \frac{e^{\alpha_i(\theta - \beta_{i3})}}{1 + e^{\alpha_i(\theta - \beta_{i3})}} \tag{17}$$

$$P(x=4) = P_4 = \frac{e^{\alpha_i(\theta-\beta_{i4})}}{1+e^{\alpha_i(\theta-\beta_{i4})}} \quad (18)$$

Then, the probability of a response of 0 is the probability of a response greater than or equal to 0 minus the probability of a response greater than or equal to 1.

$$P(x=0) = P_0 - P_1 = 1.0 - P_1 \quad (19)$$

$$P(x=1) = P_1 - P_2 \quad (20)$$

$$P(x=2) = P_2 - P_3 \quad (21)$$

$$P(x=3) = P_3 - P_4 \quad (22)$$

$$P(x=4) = P_4 \quad (23)$$

A Comparison of CTT and IRT

While both CTT and IRT approaches have been used effectively to calculate the parameters that indicate the characteristics of measurement instruments, each of them has its own advantages and weaknesses. For this part, the researcher will compare the two methods.

The basic differences between CTT and IRT include (1) assumptions and fundamental mathematical models, (2) levels of analysis, (3) parameter estimation, and (4) approaches to error(Glenn, 2005; Hambleton & Jones, 1993). CTT mainly focused on the test-level information, no accurate techniques in CTT for deciding whether test items are working according to the basic model. this makes it is difficult to predict individual item responses and to develop an instrument that assesses certain capability levels. In addition to this, CTT person and item parameters are sample and test dependent; this

makes it may be incorrect to make a comparison between test scores and generalize the estimated item parameters beyond the pilot sample for a particular test(Glenn, 2005). IRT offers solutions to these problems and provide supplements CTT by giving precise information about measurement at different ability levels. It provides both the estimation of person and item level parameters. The person parameter in IRT is latent ability level, normally embodied as theta (θ). No matter the items are dichotomous or polytomous, θ usually refers to examinees' quantity of knowledge or skill regarding the construct when the test focuses on cognitive test (for example, verbal ability) while θ presents respondents' levels of the attitude or trait in question (for example, political views) in attitude test(De Ayala, 2013) . Of course, IRT also has its own limitations. It has comparatively restricted assumptions and big sample size requirements. As a result, CTT and IRT can be used as complementary approaches to each other. Here is a table to compare CTT and IRT from the official website of University of Miami:

Table 2

Comparison of CTT and IRT

	Classical Test Theory (CTT)	Item Response Theory (IRT)
Test Length and Reliability	Longer tests are usually more reliable than shorter tests.	Shorter tests can be more reliable than longer tests.
Metric	Participants are measured on the number correct scale; items are measured on the proportion-correct scale. Score meaning is determined primarily as a location in a norm group.	Participants and items are placed on the same scale, making it possible to scale persons relative to items and vice versa.
Item Parameters	Difficulty and discrimination dependent upon the group in which they are estimated, and unbiased estimate of these parameters require a representative sample.	Difficulty and discrimination estimated in one sample from a population are linearly transformable to estimates of those parameters on another sample from the same population. Unbiased estimates do not require a representative sample.
Standard Error of Measurement (SEM)	SEM is group dependent (i.e., based on the group standard deviation) and is constant for a group, regardless of score level.	SEM is independent of the group on the measurements are taken; IRT permits calculation of a SEM for a single individual based on their performance and the item parameters. It also provides conditional SEMs that allow the SEM to vary at the different levels of latent continuum.
Trait Level Estimation (Scoring)	The number-correct score is dependent upon the number of the items in the test, the difficulty level of the items, limited to a fixed number of discrete	Maximum likelihood scoring of IRT-based tests yields a trait level estimate for an individual that is independent of (1) the number of items, (2) the difficulty level of the items, (3) the group of individuals in which the person

	values, and interpretable on a within-group normative basis	was measured, and (4) is reported on a real-number scale
Item-Model Fit	There are no precise procedures for determining whether test items are functioning according to the underlying model	Has explicit procedures for determining item fit that allow the identification of items that do not meet the assumptions of the model
Person Model Fit	Although there are procedures for determining person fit to the CTT model, they require normative comparisons.	Procedures for determining person fit allow for the identification of persons who are not well measured by the model, independently of other persons measured at the same time
Equating	Score equating requires complicated procedures that require assumptions about population score distributions.	Because persons and items are on the same scale, equating occurs automatically as a result of linking, without assumptions of score distributions. This makes it possible to compare on a common scale person measured in different groups and with different items.

Note. from Item response theory for psychologists(Embretson & Reise, 2000)

Multiple Methods

The methods of this research are to use CTT and IRT models as quantitative measures to test and refine framework of what integrative learning is, as well as using text format of qualitative data as supplementary element to build the final instrument. Downing and Haladyna (1997) proposed that both qualitative and quantitative data are required to support the use or removal of test items in developing an instrument. Based on this, this research can be considered as multiple method approaches, with qualitative and quantitative data iteratively constructing each other. Qualitative information in our process will begin with the analysis of different course proposals which can propose learning goals of each theme in the program. The proposals were submitted by the leaders of each topic, it will be regarded as qualitative data to initiate the framework. The other part of the qualitative data will be the students' answers to the open-ended questions in the survey. The researcher cycle with analysis of students' responses from the survey through a process in qualitative coding method. Therefore, developing this measurement tool were the iterations in the development process that has allowed for the study and refinement of the theoretical framework being measured, that is why this research planned two waves of data collection to produce the final instrument. This dissertation focused primarily on the quantitative methods of the framework construction and explanation, but qualitative data is also very important to finalize the final instrument.

In the documentation of the proposals of all the minor program, several core learning outcomes are identified by minor program. These include the following:

1. Discourse

2. Critical Thinking in Humanities
3. Reasoning in the Social Science
4. Reasoning in the Natural Science
5. Critique and Practice in Design and the Arts
6. Quantitative and Computational Thinking
7. Critical Analysis of Identity and Equality in the United States
8. Intercultural and Global Awareness
9. Ethical Reasoning

All the courses in the program were designed based on these core competencies. This research partly used the framework to evaluate interdisciplinary learning. Mansilla and Duraising (Mansilla & Duraising, 2007) proposed a framework to assess interdisciplinary course work. The framework includes three dimensions: disciplinary grounding, advancement through integration, and critical awareness.

In this project, I regarded critical awareness as a high-level outcome of advancement through integration. In addition to this, the author also considered "active learning" is an important dimension to the assessment. The students come from different major backgrounds, if they finish all the foundational and elective courses, they will share the same experiences in more than 18 credits, this is a very distinguished experience from a conventional general education program. Students and faculty have more chances to deeply know each other, and lead to meaningful interactions and professional relationships.

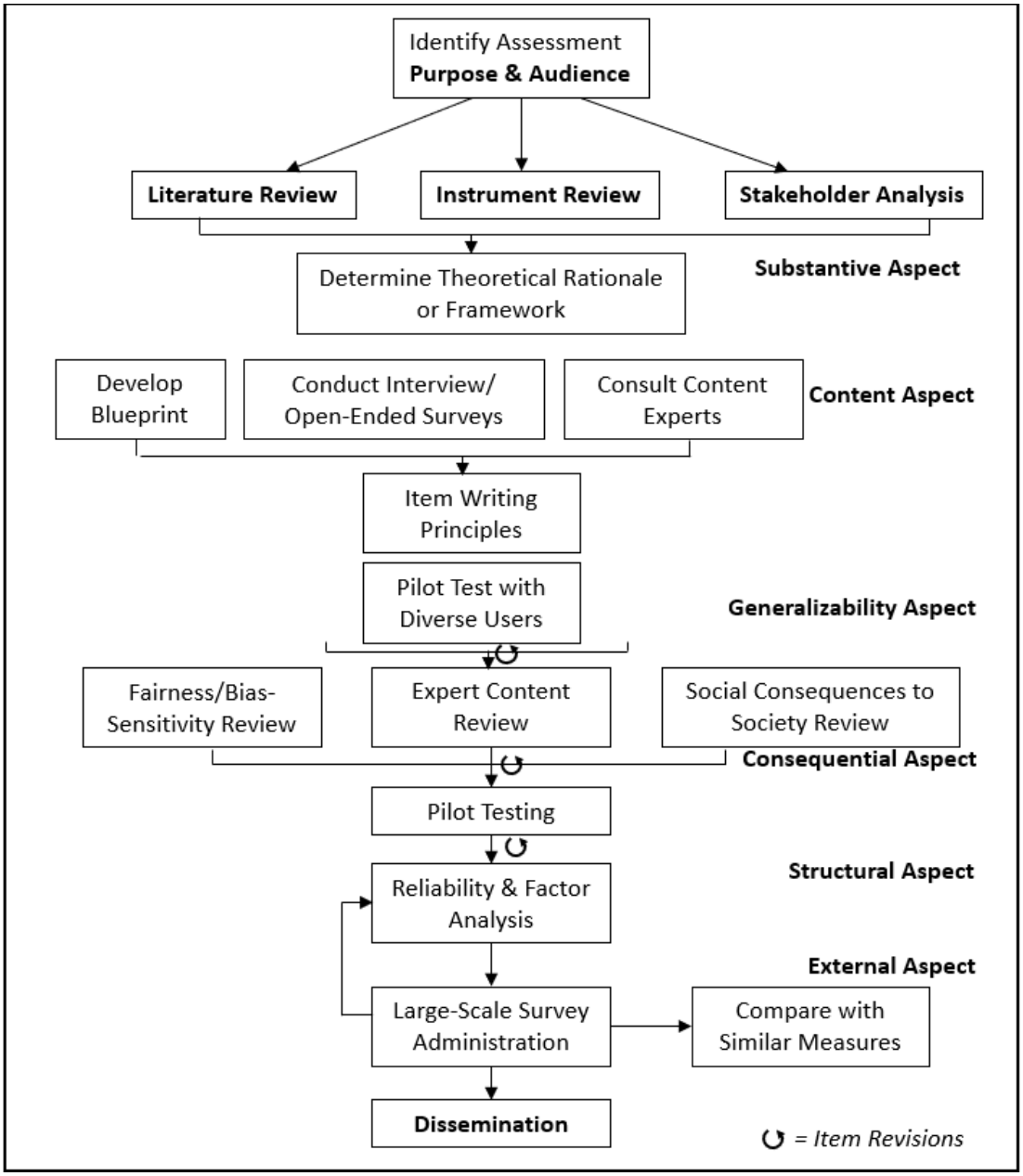
Although many studies show peer interaction can improve student learning experiences, there remains limited knowledge of where each student fit on a scale of high-level learning outcomes in an interdisciplinary program. Importantly, the main purpose of this research was not to make a comparison of student learning outcomes between students. For example, one student may have a higher level of integrative learning than another student (a norm-type comparison). This research tends to test integration at individual level based on each student. The result of this method to assess student integrative learning experiences was to develop a model that clearly identifies all the factors that might influence integrative leaning, and then use the model to decide on the dimensions necessary to measure (e.g., theme integration, peer interaction and so on). For each dimension, this research begins to define several important indicators for measurement (e.g., a theme integration scale based on the construct of the courses in each topic). Another essential part in developing the scales for each variable in the model was to design instruments to track whether the program provides support for student to learning integrative. Norm-referenced assessment ranks a student's performance against their peers in a particular cohort (K. Burton, 2006). In this research, the measurement strategy was referred to as norm-referenced measurement since we know a higher score means a higher level of integration, but it has not yet set a standard as a "good " integrative learning, only "better" integrative learning.

According to the unified theory of validity (Messick, 1998), developing an instrument includes six aspects need to be considered when constructing items (1) the quality of the content (2) consistency in responses ; (3) reliability of the observed score;

(4) extent to generalize to and across population; (5) comparisons between multimethod; (6) actual and potential use of the instrument. Purzer and Cardella converted Messick's unified theory into a diagram for development of an instrument creation.

Figure 3

Instrument Development Model: A Map based on Messick's Unified Theory of Validity(Purzer & Cardella)



Note: From Using an instrument blueprint to support the rigorous development of new surveys and assessments by Menold, J., Jablokow, K. W., Purzer, S., Ferguson, D. M., &

Ohland, M. W. (2015, June). In engineering education In 2015 ASEE Annual Conference & Exposition (pp. 26-1657).

"The gathering of supporting evidence for validating a specific test use or interpretation must begin with a careful and systematic approach to the task of creating the test items." (Downing & Haladyna, 1997). The meaning of this statement is to use a standardized approach to develop the instrument. The content creation for an instrument should start with content definition, or describing the scale of the construct domains(Haynes et al., 1995).

I mainly followed this process of frame analysis and construction. The construction of the framework started from the literature review and involved literately tacking from the data to the research literature and back again. Before the survey distribution, the program stakeholders were involved into the improvement process of the instrument. After I determined the framework, I experimented with adjusting the frame and modifying the instrument using to the data set based on the understandings of pilot study. The frame that emerged was applied to the data of field study in order to collect elements which are important in integration and could be generalized to extended cases. Eventually, the frame was further refined by applying the IRT model. As the analysis progressed, I also "thickened" the data by reading as much of the participants' open questions as possible, by adding layers of qualitative data, and by connecting the materials to the existing literature. The final modified frame is one which some aspects were validated, and others can still be modified or discarded in the future.

Towards an Initial Framework for Assessing Students' Integrative Learning

From the literature review, we get many inspirations to definite the domains from the existing instruments. We can conclude that there are many requirements to design a successful interdisciplinary curriculum for a program. This project adopted a self-report method to examine students' learning outcomes in this progressive general education program, it also borrowed and synthesized the mentioned framework and measurement before, merged the core learning outcomes conceptions into the framework, three dimensions: Integration within Minor, Advancement through Integration and Active-learning Experiences are going to be assessed.

The first dimension of the instrument is the "Integration within Minor". It will investigate the integration when student having experiences within courses. The basic is thematic integration. This sub strand will focus on the identified concepts within each discipline are consistently pertained to the interdisciplinary program theme, and whether the number of the courses is enough for the topic, the consistency between the courses, the connections between courses in different disciplines. The theme is not a loosen sampling of knowledge, the integration within the courses is the basis of this progressive general education program. This assessment method will be a way to test whether the program solves the Potpourri Problem mentioned in the previous research.

The "Disciplinary Grounding" will be taken into consideration in the instrument as the second sub strand of the dimension of "Integration within the Minor". It will assess the students' basic understanding of the knowledge they have learned within each discipline, which is the foundation of higher-level integration. Then is the integration of

knowledge in the dimension of integration within the minor, it will include the students' cognitive ability to analyze, make judgments, and take actions within the minor. This shows the comprehensive understanding of disciplinary concepts or methods which is regarded as a prerequisite for higher-level integration. It emphasizes the course content to real-world issues and the outer world related to the topic such as practical solutions that imported purposefully from different disciplines to create a workable and rational way to address a problem. The Bloom's Taxonomy is also to be considered to frame the questions in this sub-strand.

The second dimension is called "Integration beyond Minor", which shares the same meaning with the learner-initiated integration, will assess students' experiences of integration with their real world, prior experiences, major, and future life. This part of the integration will engage students more, it will assess their own construction of knowledge.

The third dimension is "Integrative Learning Environment", it will assess whether minor program provides support the students' needs in integrative learning experience. This dimension considered both faculty and peers are essential factors in the integrative learning environment.

The construction of the instrument is show as Table 3.

Core Dimension 1 - Integration within Minor. There were three sub-strands in this dimension: internal construct, disciplinary grounding and integration of knowledge. The internal construct sub-strand carefully considering the construct of the courses, whether the setting of the curriculum meet the requirements of thematic integration. The disciplinary grounding focuses on students' learning knowledge from the selected courses

and appropriateness in the use of knowledge from individual discipline. Then integration of knowledge will collect students' experiences of integrative learning in the class, whether they successfully have the cognition of breaking the boundaries between disciplines, and how the minor courses help them to establish the foundation of integration competencies.

Core Dimension 2 - Advancement through Integration. Beyond the simple internal integration, the distinctiveness of interdisciplinary learning is to generate students' ability to synthesize and integrate (Klein, 1990; Mansilla, 2005), our aim is to assess the student ability to connect knowledge taught in minor with other experiences. Minor program is different from isolated courses in traditional general education program, it digs one topic symmetrically on deep and integrative perspectives. To unbox this deep learning process gradually, this dimension is to investigate the success of connection with students' other experiences including connection to their major, connection to their prior knowledge, to know how students' experiences in the process of the linked course. Integration emphasizes on learning, it demonstrated as "ability to connect from disparate contexts and perspectives"(King et al. 2007, p.5). We borrowed some conceptions of the items in the indicator of "Reflective and Integrative Learning" in the National Survey of Student Engagement to assess the students' integration of their learning with previous experiences, and with their majors.

Core Dimension 3—Integrative Learning Environment. This dimension is to investigate whether minor provides an active learning students' engagement in minor program on how they achieve the learning outcomes. Students can learn much better by

learning actively from receiving information passively. “Active learning” is defined by researchers as learning that needs students to cognitively and meaning fully engage with the learning enviroment (Bonwell & Eison, 1991), to get involved with the information

Table 3

The Construct of the Instrument

Scales	Sub strands	Supporting Literature (Reference)
Integration within Minor (Thematic Integration& Knowledge Integration）	Internal Construct	Burton, 2001; Jacobs, 1989;
	Disciplinary Grounding	Hartesveldt & Giordan, 2008; Boix Mansilla& Dawes Duraisingh, 2007
	Integration of Knowledge	AAC&U, 2004 Borrego & Newswander, 2010; Julie Klein and Newell, 1996; Songer and Linn, 1991
Advancement through Integration (Learner-initiated integration)	Integration with Real World	Klein, 2001; NIH, 2006; NSF, 2006
	Integration with other Experiences	Duerr, 2008, p.177 AAC&U, 2004
	Integration with Future Life	Youngblood, 2007 Taylor, 2008
Integrative learning environment	Faculty-Student Interaction	Astin, 1993; Callaghan & Bower, 2012; Kim et al., 2013; Bjorklund et al., 2004
	Peer Interaction	Pascarella & Terenzini, 2005; LaPointe and Gunawardena, 2004
	Social Engagement	LaPointe and Gunawardena, 2004; Youngblood, 2007

presented, deeply thinking about it (analyzing, synthesizing, evaluating) rather than just passively receiving the knowledge(King, 1993). Thus, active learning meant students were considered to be cognitively engaged(Corno & Mandinach, 1983).

In this survey, the key indicators for this dimension are student perception of the environment climate, cooperative learning such as project- and problem-based activities, active discussion that occurs among peer students and instructors both in classroom and outside of the classroom including instructors give reflective or responsive assignments to students, the proactive reply from instructors to support student with integrative learning. Cooperative learning refers to student-centered instructional strategy that involves students both into academic and social learning(Lie, 2002). Cooperative learning is not simply assigning students to groups to do assignments. Faculty engagement is another guarantee of the course quality. In this survey, we will only investigate faculty's interaction with students.

In the literature review, we conclude that general education intends to teach students more about practical knowledge than liberal arts education. The capstone courses of different topics in Minor program mentioned, one of the essential goals is to support student career or professional development in the future. The final sub strand of Active-learning Experiences is social engagement, this sub strand will assess students' unprofessional or social interaction with faculty or other peers. Then merged these four core dimensions into the final theoretical framework.

Then comes to the scales of measures. The instrument adopted Likert scales system. In 1932, Rensis Likert published a paper called "A Technique for the

Measurement of Attitudes" and proposed the well-known Likert scale(Likert, 1932). In this historically momentous paper, he firstly mentioned the famous Likert scale, which brought dramatical impact on and widely used in the development of measurement instrument later. The Likert scale was invented to address a main limitation of other scales because Likert did not agree with the definition of attitudes as concrete quantities, rather than "a range within which responses move"(Likert, 1932, p.8). Which means, Likert believed that the attitudes type questions are more like continuous fluid and a response scale should reflect that character. Likert introduced a series of five response options that could be applied to any attitude type of items: 1 – Strongly Approve, 2 – Approve, 3 – Undecided, 4 –Disapprove, 5 – Strongly Disapprove. Likert's scale is still widely used today in various measurement instrument and survey with the more commonly used Agree substituted for Approve. Participants will select one of the response options after reading an item with their own attitude. With the development of the scale, other numbers of response options are also used; other common numbers include seven, six, and four(Jamieson, 2004). Some sort of frequency-based scale (e.g., Never, Rarely, Often, Always) can also be treated as Likert scale, because this kind of frequency is also not definite quantities but flows from lower level to higher level.

The survey develops a rubric for describing the extent to which students' sense of inquiry has developed. To avoid neutral answer, this instrument will choose 6-point Likert scale. An example of a six-point scale might be: 6 points - The students exhibit a very positive agreement of inquiry and questions. 5 points – The students exhibit a positive agreement of inquiry and questions. 4 points- The students exhibit somewhat

positive agreement of inquiry and questions. 3 points- The students exhibit somewhat negative agreement of inquiry and questions. 2 points - The students exhibit most part negative agreement of inquiry and questions 1 point - The student has not exhibited any sense of inquiry.

For the “Integration within Minor” and “Integration beyond Minor” dimensions, students rated all items on a 6-point Likert-format scale with descriptors at each point (1 = Strongly Agree, 2 = Agree, 3 = Somewhat agree, 4 = Somewhat disagree, 5 = Disagree, 6 = Strongly Disagree)., the questions in “Integrative Learning Environment” were set as the Likert Scale of four-point (1 = Many, 2 = Some, 3 = A few, 4 = None). When analyzing data, the researcher will reserve the order to (1 = Strongly disagree, 2 = Disagree, 3 = Somewhat disagree, 4 = Somewhat agree, 5 = Agree, 6 = Strongly agree) and (1 = None, 2 = A few, 3 = Some, 4 = Many).

Table 4 shows the meaning of different scores in the construct of the instrument. From 1 point very unsupportive to 6 point very supportive. Each score has its own different meaning in different scale. It indicates the levels of students’ perceptions in different dimensions of integration. For example, 1 point in the internal construct means “The minor courses are not isolated with each other: all the courses are isolated with each other, there is no connections between different courses, can’t see any integrative knowledge between courses”, it is the narrative for the different scores that participants choose.

Table 4

Scales of Measures

Variables	Sub strands	Levels
Integration within Minor	Internal Construct Disciplinary Grounding Integration of Knowledge	• The minor courses are not isolated with each other: all the courses are isolated with each other, there is no connections between different courses, can't see any integrative knowledge between courses. • The minor courses are not very connected with each other: the courses are not very connected with each other, the integrative knowledge between courses is not very much. • The minor courses relate to each other: all the courses relate to each other, there are some connections between different courses, student can learn integrative knowledge between courses. • The minor courses are closely related to each other: all the courses are well connected with each other to show the theme of the minor. There are many connections between different courses, student can deeply learn integrative knowledge between courses.
Integration beyond Minor	Integration with Real World Integration with other Experiences Integration with Future Life	• The program does not help the integrative learning at all: minor program does not help student to understand the real-world issues at all, does not related to my previous experiences, and has no help for the future career and professional development. • The program does not help the integrative learning: minor program does not help student to understand the real-world issues that much, does not related to my previous experiences very much, and has seldom help for the future career and professional development. • The program helps the integrative learning: minor program sometimes helps me to understand the real-world issues, sometimes related to my previous experiences, and has some help for the future career and professional development • The program helps the integrative learning a lot: minor program really helps student to understand the real-world issues, it closely related to my previous experiences, and helps student to make the road for my future career and professional development

Integrative learning environment	Faculty-Student Interaction	• Very Unsupportive: the program environment is not supportive at all, almost not the interaction between students and instructors. It does not promote integrating knowledge learned in class and does not apply the knowledge to the real-world environment and issues. It does not provide opportunities to know instructors and peers at all. • Unsupportive: the program environment is not very supportive, the interaction between students and instructors are not frequent and effective. It does not promote integrating knowledge learned in class and applied the knowledge to the real-world environment and issues. It does not provide opportunities to know instructors and peers that might be useful in my future career. • Supportive: the program environment is supportive, the interaction between students and instructors are good. It promotes integrating knowledge learned in class and applied the knowledge to the real-world environment and issues. It provides opportunities to know instructors and peers that might be useful in my future career. • Very supportive: the program environment is very supportive, the interaction between students and instructors are frequent and effective. It effectively promotes integrating knowledge learned in class and applied the knowledge to the real-world environment and issues. It helps student to establish good relationship with the instructors and peers that might be useful in my future career.

Item Development

The purpose of this instrument is to evaluate students' integrative learning in a general education program, according to the frame, the whole instrument is composed of three sets of related items— Integration within Minor, Advancement through Integration and Integrative Learning Environment. Integration within Minor, Advancement through Integration items are 1-6 Likert questions that ask about students' perceptions of their advancement of integration in the program. There are 24 items and two open ended questions in the first version of instrument, 25 items and three open ended questions in the second version of the instrument (Does not include the demographic information and other informative items). Responses to the quantitative items are used to predict the probability that students' perceptions to endorse an answer. The format of the instrument is meant to imitate a context in which students are asked to answer their experiences in the program. The instrument was administered online, and it contains some item sets that are not in the dimensions of the framework such as demographic information. A sample item that is not in the framework is given in Figure 3-4. The results from this instrument are to be used in program level.

Table 5

Demographic information Item

Demographic Items	Content
D1	What is your name (last name, first name)?
D2	What is your school email address (e.g., name@**.edu)?
D3	Your sex
D4	Your race/ethnicity
D5	What is your primary major?
D6	What is your secondary major (if applicable)?
D7	Your class standing

Table 6

Figure Item Sample in Core Dimension

Scales	**Sub strands**	**Items**
Integration within Minor(Thematic Integration& Knowledge Integration）	Internal Construct	I feel that the number of classes I completed for this minor was sufficient to meet the goals of the minor.
	Disciplinary Grounding	I understand the concepts taught in the minor.
	Integration of Knowledge	Over the course of the minor, I applied knowledge across different assignments.
Advancement through Integration (Learner-initiated integration)	Integration with Real World	I have identified problems or issues related to this minor in the real world.
	Integration with other Experiences	I have made connections between things I learned in this minor and things I already knew.
	Integration with Future Life	I have made connections between things I learned in this minor to my future career.
Integrative learning environment	Faculty-Student Interaction	Over the course of the minor, how many times did you interact with the instructor as part of this minor?
	Peer Interaction	Over the course of the minor, how many times did you discuss your ideas with classmates?

	Social Engagement	During the course of the minor, how many times did you interact with other students outside of class?

Here is the table of item number for each scale and sub strand in Spring 2021, gives a total of 25 quantitative items.

Table 7

Number of Items of Content by Operation in Spring 2021:

Scales	**Sub strands**	**Items Number**
Integration within Minor (Thematic Integration& Knowledge Integration）	Internal Construct (IC)	3
	Disciplinary Grounding (DG)	4
	Integration of Knowledge (IK)	4
Advancement through Integration (Learner-initiated integration)	Integration with Real World (IR)	2
	Integration with other Experiences (IE)	2
	Integration with Future Life (IF)	1
Integrative learning environment	Faculty-Student Interaction (FSI)	4
	Peer Interaction (PI)	4
	Environment Climate (EC)	1

Instrument Evaluation

Item evaluation consists of four stages. The instrument was managed on an online survey platform Qualtrics—an online survey platform. The evaluation from every stage were used to revise the items before the next step began. In the first state, the team who has the researcher of this dissertation, the director of the program who is also a PhD from education area, another PhD student and who is also a staff in the program reviewed the

items for wording, clarity of conceptions. The next stage involved staffs who are working in the program and undergraduate students who had completed the program. Stage 3 was the pilot testing; 39 undergraduate students completed the assessment instrument. These data were used to refine items to the next version of the assessment. In the final step, 151 students completed the updated version of the instrument. Data from the Stage 3 and 4 were combined for the final analyses.

This research is a part of the assistantship, to provide the reports leaders in minor, I submitted an IRB application on December 10, 2019 and has been approved in Feb 10, 2020 for purposes of both work at the university and for the dissertation (see Appendix). In the IRB, I mentioned the quantitative and qualitative methods I will use and presented that the project will be utilized as my doctoral thesis. The IRB department replied since this was program evaluation, it was beyond the scope of federal guidelines for human subject research, its protocol approved as "Not Human Subjects Research". Before the prospectus exam, I wrote an email to IRB office to introduce the research I had done at that point, to confirm whether the project required any amendment for IRB, they replied there was no need to do any revision.

Stages One: Experts Review. In Stage 1, all the items were evaluated by experts for item quality. The experts were stakeholders including director of the program. The director is the assistant provost at the university and closely work with all instructors and students in the program. He is a PhD graduated from an Educational Psychology program and responsible for the whole program. The other one a PhD student who also working as a staff in the program, he had a M.A degree in statistics. They both reviewed the items for

discussion the construct of the wording, clarity of conceptions in the frame. To improve the item quality in this stage, we document comments and data on the items through online platform. We also scheduled 6 meetings to reword and refine each item.

Stages Two: Internal Think. After developing the first version of the instrument was finished, the researchers presented the instrument to the Office of Undergraduate Education at the University. The staffs who working in the program also give some constructive suggestions on the items. To test the instrument, this involved 3 staffs who are working in the program and 2 undergraduate students who had just completed the program. They completed the online assessment while expressing their thoughts about each item. Since participants did not have any research background, the data and feedback gathered during stage 2 were analyzed and used to revise the verbality of the items for better clarity. The researcher sent instrument link through emails; the participants accessed items through a Qualtrics survey on the researcher's account. In this stage, the participants would be asked to write down their comments on the instrument.

Stage Three: Pilot Study. The researcher planned to implement the pilot study in Spring 2020, but because the COVID-19 pandemic started in the spring semester, the whole campus was in a chaos, every student was trying their best to get used to the online courses and their interrupted lives by virus. At that time, the researcher believed the pandemic would over soon, to make research more reliable regarding the integrative learning dimension in the instrument, the survey distribution plan was postponed to fall, 2020. In the Fall 2020, there was no evidence that the pandemic would end. The pilot study had to be implemented in the fall semester. The pilot data was collected from

October 2020 to December 2020, the data collection of field test was applied in the following semester and completed in May 2021. From the August 2020, the researcher started to develop the assessment instrument for the pilot study. After several rounds of discussion and revision, the instrument for the pilot study was submitted to the researcher's academic committee members in the preliminary exam. The academic committee members gave the researcher many good suggestions. The first version of instrument was finalized in October. Before the distribution of the survey, the researcher invited several colleagues and students who were working in the Office of Undergraduate Education to do a small survey testing. The participants gave me some suggestions such as "there are too many overlap questions", "some questions are not very clearly to me". Their feedbacks helped me to finalize the wording of the survey. From October 2020, the director of General Education at the University leaded the connections between with the researcher and leaders of the minors which had a capstone course in Fall 2020. The researcher suggested instructors to distribute the survey as a credit-bear assignment, but since it was in the middle of the semester, some instructors refused to do so but promised to encourage students to participate. Finally, five capstone courses: Philosophy, Politics, and Economics, Disabilities Studies, Global Food Security and Health, Organizational Leadership and Civic Agriculture and Food Systems participated in the pilot study. Before the distribution of the survey, two students who already graduated from minor program did a testing response for the instrument, to make sure all the words and sentences could be understood by common participants.

The survey was distributed through Qualtrics link by the instructors to students, the researcher also did a PowerPoint slides flyer and posted on canvas from late October 2020 to middle December 2020. During the data collection process, two reminder emails and a promotion video recorded by the researcher were sent to instructor to incent student to participate.

Stage Four: Field Test. Based on the analysis result of the pilot study and deeper thinking about the instrument, the researcher discussed with the stakeholder again and revised some items in the first version of the instrument. The revision happened in March 2021. Since the sample of the pilot study is too small, although indicators of some items didn't achieve the statistical standards, the researcher decided to keep them in the second wave of data collection regarding their importance for the integrative learning.

The timeline of the data collection of field test wave one was set in Spring 2021. During the pilot study, because time point of the instructors got the information of the assessment was in the middle of the semester, they refused to use assignment-bear incentive. These resulted in a low response rate. Based on these experiences that were not good enough, to gather more supports from instructors to incent students to participate in the survey, the social connection emails were sent back and forth to the leaders of the minors which had capstone courses before the start of the semester, in middle January. These emails informed them the purposes of the program evaluation and got leaders' permission to collect data. Then the researcher sent the instruction to the instructor of the capstone courses, to make them be acquainted with the evaluation process before sending the survey to students. Finally, 16 capstone courses: Blue Planet, Community Systems

and Engagement, Data and Decisions, Disabilities Studies, 'Ecological Cities, Event and Experiences Management, Innovation, Integrated Security, 'Language Sciences, Organizational Leadership, Philosophy, Politics and Economics, Science, Technology & Law, Strategic Communication, Sustainability, Visual Arts and Society and Adaptive Brain and Behavior participated in the second wave of data collection. The survey was also distributed through Qualtrics link by the instructors to students on the class PowerPoint slides and canvas.

The formal data collection was started from April 7 and ended in the middle May 2021. One reminder email and promotion video recorded by researcher were sent to instructors to enhance the response rate in the process of the assessment.

The second wave of data was collected in spring 2022, since the researcher started to work at other university, the data management was transferred to the office. Before the distribution of the survey, the researcher initiated a meeting with the director of the program and the graduate assistant. According to the analysis result of the data, the researcher proposed suggestions to revise the instrument. The graduate assistant who is the PhD candidate also shared her opinions with the researcher on the google doc, for example the interpretation with some items. After all the revision and discussion, the survey was distributed through a new platform called QuestionPro. The survey was closed on May 16, 2022.

Sample Selection and Data Collection

During 2019 May to 2021 February, the researcher was working in the Office of Undergraduate Education at the University as a graduate assistant who was responsible

for supporting the whole minor program. In this process, the researcher was acknowledging much information of the program. Based on the work, the researcher got opportunities to meet and talk with some of the minor leaders and instructors, and introduced the project goals to them, made them understand the essence of the project and exchange their ideas and suggestions on the project.

After that, the office decided to collect data from the 2020 spring semester. But the COVID 19 was starting to spread in 2020 March, the classes started transfer from onsite to online. That was a hard time for all student, everyone was trying to find a best way to adjust their study and living method. Under this background, the researcher and the office decide to postpone the data collection to 2020 fall. Since the majority of capstone courses are in spring, finally the sample of this study were the students in the capstone courses of the program in 2020 Fall and 2021 Spring at this public land granted university.

The researcher plans to only distribute the instrument to the students in the capstone courses. Since the researcher can't get the enrollment information whether students are from major and minor, the research did not have the enrollment data, so the population number remains unknown. This is the reason why we can Several reasons made the researcher to only chose the students in the capstone courses as the research sample. The capstone courses were applied, project-based, which meant students had to bring all the knowledge that they had learned through minor together. The students who were in the capstone courses had a comparative completed experiences about minor program. They were the best target groups to give a fair answer to the instrument that aim

to know their learning experiences. From the literature review, diversity was an important factor that could affect integrative learning., each minor topic included several elected courses but only one capstone course. It meant the students from different backgrounds who enrolled in one topic would merge into the same capstone course and shared the same course experiences. Finally, the capstone courses specially emphasized the cooperative learning during the instructional process which was also an important dimension of student integrative learning experience this study tried to investigate. In addition to that, all the students in capstone courses had already completed other levels courses. They experienced not only the online capstone course during the pandemic, but also face to face courses in minor. This would reduce the impacts that pandemic brought to their learning experiences. Finally, the researcher was working as a graduate assistant at the Office of Undergraduate Education when initiating the idea of this research and developing the instrument. The office is the center to administrate and support the daily operation of Minor program. Because of the special job position, the researcher could conveniently make social connections to the instructors, leaders of each minor, and have direct dialogues with them when collecting data. This improved much of the possibility to conduct success research.

The researcher listed all the minors and listed all the capstone courses, then looked up the class table on the official website of the university. Five topics in the minor program that has capstone courses in 2020 fall semesters. All the minor students in the classes are the targeted sample for the study. Then the researcher sent emails to ask the instructors whether they would like to participate in the research. After their “Yes”

answer, then the researcher will send each class an individual link to the survey. The instructors will send the link to the students only from the minor and share the link on canvas for students to answer the survey.

To improve the response rate, the researcher also did a PowerPoint slides flyer and posted on canvas from late October 2020 to middle December 2020 and April 2021 to middle May 2021. During the data collection process, two reminder emails and a promotion video recorded by the researcher were sent to instructor to promote student to participate in the pivot study and first wave of field study.

The second wave of field study was collected and administrated by the Office of Undergraduate Education in Spring 2022.

Initial Data Analysis

Item analysis is a procedure for quantifying various characteristics of test items (Meyer, pp. 40). This research utilized SPSS 27.0, jMetrik and IRTPRO 4.2 student version to prepare and analyze the data.

When we interpret a score of an individual by comparing his score to those of other individuals is norm-referencing. When we interpret a person's performance by comparing it to some pre-specified standard or criterion of proficiency in the process of criterion referencing. In a criterion-referenced test, the examinee scores indicate the degree of achievement in the field of the content. The interpretation of the score pivots on the clear definition of the content in the instrument and the alignment of the test to the field. According to the goal of this project, which it to know whether minor program

meets its requirements of integration, this research should be a criterion referencing assessment.

Pilot Study

The data cleaning was completed by SPSS 27.0. The cleaning process included changing the name of the variables according to the framework, deleting the duplicates that Qualtrics recorded, deleting the students who completed under 80% of the survey. After cleaning the data, the total number of participants for the pilot study was 39, the sample is too small to do the item analysis in Jmetrik. So, the researcher decided to continue use SPSS 27.0 to do basic item analysis of the data, this process could obtain the basis descriptive information of the data, and also provide some information for the future instrument improvement.

The capstone class is only included in the curriculum of all minor programs, but students in the capstone course are not only from minor. For example, the capstone courses for Organizational Leadership minor is MGT 4354, this is not only capstone course for minor, but also core course for some major in business school. The students in this course are coming from both Organizational Leadership minor and other major. The students who come from major did not participate in the project. Since we did not know which students was from major or minor, response rates for students, therefore, were not calculated because they would be inaccurate and misleading.

Table 8

Participating Capstone Courses Ordered by the Number of Students in Each Course

Course topic	Capstone course	N

Civic Agriculture and Food Systems	ALS 4214	6
Disabilities Studies	HD4714	17
Global Food Security and Health	ALS 4244	7
Organizational Leadership	MGT 4354	5
Philosophy, Politics and Economics	PHIL/PSCI/ECON 4884	4
Total		39

Five classes participated in the pilot study. The table above listed all the five capstone courses that participated in the evaluation. In the data of pilot study, female students took 82.1% of the participants, male students took 17.9% of the participants. The percentage of female students was much greater than male students.

Table 9

Demographic Information of Students in Pilot Study

Question	Response	N
Sex	Male	7 (17.9%)
	Female	32 (82.1%)
	Total	39 (100%)
Race/Ethnicity	Asian or Pacific Islander	2 (5.1%)
	Hispanic	2 (5.1%)
	White or Caucasian (not Hispanic)	32 (82.1%)
	More than one of these	3 (7.7%)
	Total	39 (100%)
Class Standing	Freshmen	0 (0%)
	Sophomore	1 (2.6%)
	Junior	6 (15.4%)
	Senior	32 (82.1%)
	Total	39 (100%)

The white students were still the main body of the participants, it took 82.1% of the participants. The Asian and Pacific, who took 5.1% of the population. 7.7% of the

participants who reported their race more than one of the races. Black or African and Hispanic were both 5.1%. It is normal that 32 senior students took 82.1% of the participants, because the participants groups were capstone courses students, they already completed other levels of courses, 6 students where Junior, 1 was sophomore.

Here is the table of the reliability for three dimensions:

Table 10

The Reliability of Different Dimensions in the Instrument

	Cronbach's Alpha	N of Items
Integration within Minor	.867	11
Integration Beyond Minor	.593	5
Integrative learning environment	.595	8

The Cronbach's Alpha for the Integration within Minor was 0.867 which is considered a strong indicator of reliability. The reliability of Integration Beyond Minor was 0.593, and the reliability of Integrative learning environment this factor was 0.595, were unexpectedly low. However, given that this is a pilot study, and the sample was very small (n=39), additional analysis was required for consideration before a decision was made about the items on the instrument.

Table 11

Decision of the Items in the Pilot Study

Item Description	Problem	Decision and Reason
During the course of the minor, how many times did you interact with other students outside of class?	Misinterpreted	Revise to more general direction

Over the course of the minor, how many times were you asked to reflect on, revise, and resubmit an assignment?	Negative correlation	Left- reflection is an important factor
Over the course of the minor, how many times did you use digital tools(e.g. Canvas, Google docs, discussion boards, etc.) to work collaboratively?	Leaning response	Revise to more general direction

The items which had low or negative correlations with other items in their dimension were listed in Table 11 along with the decisions whether to keep or drop them. After discussion with advisor and experts, we all thought the results which can be resulted by the small number of the sample of pilot study, the researcher decided not to delete the items but revised some of them at this stage.

Field Test

Spring 2021. This instrument changed "Over the course of the minor, how many times did you use digital tools (e.g. Canvas, Google docs, discussion boards, etc.) to work collaboratively?" to "Over the course of the minor, how many times did you work collaboratively?". Because for all the undergraduate students who enrolled courses at the university had to use canvas to follow course updates. They would be most likely to choose "many" as their answer. To avoid this problem, the item was revised into a more general direction as "Over the course of the minor, how many times did you work collaboratively?". Also, "During the course of the minor, how many times did you interact with other students outside of class?" was updated to "During the course of the minor, how many times did you interact with other students enrolled in the minor outside

of class?" This revision limited the interaction between students only in the context of the minor, but not other programs which may happened to some students at the same time.

To get more information on integrative learning to finalize the framework for the instrument, the researcher add "Please give one or two examples of times you made connections between courses in this minor.", "Please give one or two examples of times when you connected experiences in this minor with your experiences in the real world.", "Please give one or two examples how you interact with the people(faculty, students) in this minor." to the instrument.

To make the instrument focus more on integration, after discussing with the director of the program and delete two open-ended questions, the researcher deleted "What are 2-3 of your most positive takeaways from completing the minor?", "What are 2-3 ways you think the minor could be improved?", because these two questions were too broad to get the effective information on integrative learning.

151 students participated in this round of data collection. below is a table of the reliability of the whole instrument. The Cronbach's Alpha for the three dimensions increased in this wave, the first and second dimensions were over 0.8, which means the internal construct were very reliable.

Table 12

The Reliability of Different Dimensions in the Instrument

	Cronbach's Alpha	N of Items
Integration within Minor	0.931	11
Integration Beyond Minor	0.828	5
Integrative learning environment	0.704	9

Spring 2022 The researcher did not participate in the spring 2022 data collection, but the office gave permission to the researcher to use the data. The second wave of field study started in early April and closed on May 16th.

Because the timeline of this wave of data collection was after the qualitative analysis of the pilot study and the first wave of the data, the researcher revised the instrument according to the analysis result. One item was added to the construct course called THEME4: "The classes I took early on in the minor prepared me well for the classes I took later on in the minor." One item was added to the integration of learning: "I have reflected on the knowledge I have learned throughout this minor." One item was removed because it was considered duplicative: "I have made connections between things I learned in this minor and my major." Finally, one item was added: "Over the course of the minor, how many times did you work collaboratively?".

Table 13

Decision of the Items in Spring 2021

Item Description	Problem	Decision and Reason
The classes I took early on in the minor prepared me well for the classes I took later on in the minor		Added -To test the vertical construct of courses
I have reflected on the knowledge I have learned throughout this minor		Self-reflection is very important
I have made connections between things I learned in this minor and my major.	Overlap question	Deleted

Over the course of the minor, how many times did you work collaboratively	Overlap question	Deleted

Table 14

Number of Items of Content by Operation in Spring 2022:

Scales	**Sub strands**	**Items Number**
Integration within Minor (Thematic Integration& Knowledge Integration）	Internal Construct (IC)	4
	Disciplinary Grounding (DG)	4
Advancement through Integration (Learner-initiated integration)	Integration of Knowledge (IK)	5
	Integration with Real World (IR)	2
	Integration with other Experiences (IE)	1
	Integration with Future Life (IF)	1
Integrative learning environment	Innovation	1
	Faculty-Student Interaction (FSI)	4
	Peer Interaction (PI)	2
	Environment Climate (EC)	1

Students completed the initial instrument of Spring 2021 (Appendix D) and the updated version of the instrument in Spring 2022 (Appendix E). As shown in the appendix, the final version of the instrument from analysis of the data that resulted from the pilot test.

Reasons to Combine Field Study Data

The two waves of field study data cleaning were all completed in SPSS. To get more statistical meaning from the data, the researcher merged the two waves of data to

achieve a larger sample size. Evidence is provided below to support merging two waves of data.

Demographic information. For spring 2021 data, the researcher selected the finished variable as “finished”, selected the progress variable ad “100” in Qualtrics which means the percentage of instrument that the respondents have completed as 100%. Then SPSS would select students who completed the survey. Then used SPSS to do the descriptive analysis of demographic information. Then the researchers checked duplicates, all the duplicates had been removed. The survey didn’t asked questions of minor because each minor has its own link, the researcher added minor name to each dataset in SPSS by syntax, then combined them together to the final dataset.

From the table below, female takes most of the respondents, the majority of the respondents are white or Caucasian, the following race is Asian or Pacific Islander, for the class standing, Junior and Senior are the main body of the participants. Here is the table for the course information of respondents. In this round of survey, the respondents were from 15 capstone courses.

The data in Spring 2022 was collected on the platform QuestionPro because the university no longer supported Qualtrics. Except the items on the instrument remained the same. The variables “Finished”, “Status” and “Progress” particularly designed by Qualtrics to clean data were removed. The data was collected and managed by another graduate assistant. After closing the survey, the graduate assistant shared the raw data with the researcher. Through raw data, the researcher can only delete the cases by detecting missing data and duplicates. The SPSS was still used in this round of

descriptive data analysis. Although the number of respondents were different from the wave one field study, the female, white, junior and senior students are still the majority of the respondents, which means the demographic information of the two waves data of field study hasn't changed that much.

Table 15

The Demographic Information of Participants in Spring 2021

Question	Response	N
Sex	Male	40(26.5%)
	Female	110 (72.8%)
	Missing	1(0.6%)
	Total	151 (100%)
Race/Ethnicity	Asian or Pacific Islander	23 (15.2%)
	Hispanic	6 (4.0%)
	Black or African American	7(4.6%)
Race/Ethnicity	White or Caucasian (not Hispanic)	103(68.2%)
	More than one of these	10 (6.6%)
	Another race/ethnicity not provided here	1(0.7%)
	Missing	1(0.6%)
	Total	151 (100%)
Class Standing	Freshmen	3 (2.0%)
	Sophomore	14(9.3%)
	Junior	34(22.5%)
	Senior	99(65.6%)
	Missing	1(0.6%)
	Total	151 (100%)

Table 16

The Course Information of Participants in Spring 2021

Course topic	Capstone course	Number
Innovation	MGT/ENGE/IDS 4094	13

Ecological Cities	FREC/BSE/HORT/LAR/SPIA 4554	14
Integrated Security	PSCI/BIT/CS 4164	6
Event and Experience Management	HTM 4434	12
Blue Planet	ALS/WATR 4614	3
Science, Technology & Law	STL 4314	5
Strategic Communication	COMM 4404	5
Sustainability	GEOG/NR 4444	12
Visual Arts and Society	ART 4104	27
Language Sciences	ENGL 4084	11
Data and Decisions	BIT/MGT 4854	18
Disabilities Studies	HD 4714	13
Adaptive Brain and Behavior	HD 4714	1
Organizational Leadership	MGT 4354	6
Philosophy, Politics and Economics	PHIL/PSCI/ECON 4884	6
Total		151

What's more, there was not any systematic differences between the 2021 and 2022 student populations since the process of recruiting students didn't change much. It strengthens that this research can combine the two waves of data. The researcher did not participate the data collection process. The survey has one question asked, "What is the Pathways minor you are completing this survey for?". There were 75 in 77 respondents answered the question. The table below is showing the response for minor. There were 13 minors participated in this survey including one new minor called Global Business Practices to Improve the Human Condition.

From the descriptive tables showed below, we concluded that the demographic information did not change much between the two waves of data.

Table 17

Demographic Information of Participants in Spring 2022

Question	Response	N
Sex	Male	23(29.9%)
	Female	53(68.8%)
	Prefer not to answer	1(1.3%)
	Total	77 (100%)
Race/Ethnicity	Asian or Pacific Islander	5 (6.5%)
	Hispanic	3 (3.9%)
	Black or African American	5(6.5%)
	White or Caucasian (not Hispanic)	63(68.2%)
	More than one of these	10 (6.6%)
	Another race/ethnicity not provided here	1(0.7%)
	Total	77(100%)
Class Standing	Freshmen	0 (2.0%)
	Sophomore	0(9.3%)
	Junior	7(11.7%)
	Senior	68(88.3%)
	Total	77(100%)

Table 18

Demographic Information of Participants in Spring 2022

Course topic	Capstone course	Number
Innovation	MGT/ENGE/IDS 4094	13
Biodiversity Conservation	FIW4114	6
Civic Agriculture and Food Systems	ALS 4214	2
Event and Experience Management	HTM 4434	7
Blue Planet	ALS/WATR 4614	4
Science, Technology, & Law	STL 4314	2
Strategic Communication	COMM 4404	10
Global Business Practices to Improve the Human Condition		12
Language Sciences	ENGL 4084	6
Data and Decisions	BIT/MGT 4854	6

Adaptive Brain and Behavior	HD 4714	5
Organizational Leadership	MGT 4354	14
Philosophy, Politics and Economics	PHIL/PSCI/ECON 4884	4
Total		75

Comparison of Means. To take a closer look at the data from statistical perspective, the researcher grouped common items which appeared in two waves into the categories “Integration within Minor”, “Integration beyond Minor” and “Integrative Learning Environment” according to the framework, then calculated the numerical mean of each dimension by SPSS.

Table 19

Descriptive Data of Common Items in Three Dimensions

	Wave	N	Mean	Std. Deviation	Std. Error Mean
Common Integration within Minor	1	148	5.2690	.70153	.05767
	2	79	5.2635	.62113	.06988
Common Integration beyond Minor	1	149	5.0604	.80258	.06575
	2	79	5.1930	.70136	.07891
Common Integrative Learning Environment	1	149	3.4438	.59122	.04843
	2	79	2.6123	.36831	.04144

The research used SPSS to do two sample t-test, the t-test result is showing below. This table is showing the mean of different dimensions in two waves of data. Although the numbers of participants in two waves field study are quite different, from the raw means, we can see only the mean of the third dimension are showing comparatively big.

Table 20

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means			
						Significance	
		F	Sig.	t	df	One-Sided p	Two-Sided p
Common Integration within Minor	Equal variances assumed	.100	.752	.059	225	.477	.953
	Equal variances not assumed			.061	176.874	.476	.951
Common Integration beyond Minor	Equal variances assumed	.058	.809	-1.239	226	.108	.217
	Equal variances not assumed			-1.291	178.547	.099	.198
Integrative Learning Environment	Equal variances assumed	15.727	.000	11.377	226	.000	.000
	Equal variances not assumed			13.044	220.151	.000	.000

Table 20 Continued

Independent Samples Test

t-test for Equality of Means				
			95% Confidence Interval of the Difference	
	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	.00552	.09402	-.17975	.19079
Equal variances not assumed	.00552	.09060	-.17328	.18432
Equal variances assumed	-.13264	.10705	-.34357	.07830
Equal variances not assumed	-.13264	.10271	-.33532	.07005
Equal variances assumed	.83145	.07308	.68745	.97546
Equal variances not assumed	.83145	.06374	.70583	.95707

The two-sample t-test was suggesting the same conclusion with the descriptive output of the raw data. The p-values of the first two dimensions were greater than 0.05, which indicated the means are not significantly different. But the p-value of the third dimension “Integrative Learning Environment” was smaller than 0.05, this meant the means of the third dimension in the two waves of data had statistically discrepancy.

Table 21

Independent Samples Effect Sizes

				95% Confidence Interval	
		Standardizer[a]	Point Estimate	Lower	Upper
Common Integration within Minor	Cohen's d	.67474	.008	-.265	.281
	Hedges' correction	.67700	.008	-.264	.280
	Glass's delta	.62113	.009	-.264	.282
Common Integration beyond Minor	Cohen's d	.76915	-.172	-.445	.101
	Hedges' correction	.77172	-.172	-.444	.101
	Glass's delta	.70136	-.189	-.463	.086
Common Integrative Learning Environment	Cohen's d	.52509	1.583	1.273	1.891
	Hedges' correction	.52684	1.578	1.268	1.885
	Glass's delta	.36831	2.257	1.808	2.701

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

Statistical significance does not imply practical significance, because p-values strongly depend on sample sizes. To understand the t-test result better, the researcher decided to take the effect size values (Cohen’s d) of all dimensions as a reference. All the

effect size values were over 0.5 but don't reach 0.8, which indicated the difference of the two waves data is a medium effect. Meanwhile, the third dimension is the smallest effect size, although the p-value is smaller than 0.05, which means the two-sample test of this dimension is not big enough.

The two sample t test results reveals that the means of three dimensions are not firmly differentiated from the statistical testing, which means that two waves of data can be combined in the later analysis.

Exploratory Factor Analysis Loading. Then the researcher used SPSS to do Exploratory Factor Analysis of the common items.

Table 22

Rotated Component Matrix[a]

	Wave 1			Wave 2		
	Component					
	1	2	3	1	2	3
THEME1	.805			.799		
THEME2	.695			.652		
THEME3	.737			.722		
DISGR1	.732			.652		
DISGR2	.796			.823		
DISGR3	.813			.763		
DISGR4	.644			.695		
KNOWLEDGE1	.754			.670		
KNOWLEDGE2	.678			.726		
KNOWLEDGE3	.659			.776		
KNOWLEDGE4	.637			.795		
BEYMIN1		.782		.583	.548	
BEYMIN2		.829		.514		
BEYMIN3	.475	.607		.607		
BEYMIN5	.	.666		.638	.499	
LE1	.693			-.424	-.478	
LE2	.700			-.455	-.477	
LE3			.516		.703	

LE5	.603	.708	
LE6	.634	.648	
LE4	.643	.575	.422
LE8	.570		.836
LE9	.632		.697

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 6 iterations.

The loadings of the common items for each factor are quite uniform for the first dimension in two waves, they all loaded on the first factor. In the second dimension, four items all loaded on second factor two in the wave one while there were only two loaded on that factor in the second wave. For the dimension 3, there were two items loading on the first factor, six items loading on the third factor in the first wave of field study, however in the second wave, there were two items loading on both first and second factors, three items loaded on the second factor, two loaded on the third factor, one loaded on both second and third factors.

The EFA loading showed that in 23 items, 18 items in the two waves loaded on the same factors. This result indicated that two waves of data can be combined in the item analysis.

Plan of Quantitative Data Analysis

Based on the previous analysis, the wave one and wave two of field study will be combined in the later analysis. All the items will be modelled using Classical Test Theory (CTT) and Item Response Theory (IRT). The items of the dataset were subject to item analysis, applying standard statistical procedures. The research first examined the reliability of the whole instrument and reliability of each dimension. and then conducted

a simple item analysis through CTT and Rasch model. The index of discrimination was obtained by calculating the corrected item-total correlation coefficients (r) for each item with its hypothetical scale. After that, the analysis tested the IRT model assumptions of unidimensionality which was done using confirmatory factor analysis (CFA), the model fit indicators such as GFI, RMR, and RMSEA, as well as the percentage of variance that the eigen gave the insights of whether the dimension satisfied the pre-assumption of IRT. The next step was comparing the polytomous model fits including partial credit model, graded response model, general partial credit model and nominal response model. Based on the main indicator of AIC and BIC, chose the best model to estimate item parameters. Based on the selected model of IRT, the researcher will do the item analysis. Through the analysis of parameter evaluation, total information curve, trace lines and item information curves to see the quality of the items. Finally, the researcher will propose the revision of the instrument according to the analysis result.

The analysis process will be done through jMetrik and IRT pro. The Chapter 4 will describe the results and findings of all the steps. The Chapter 5 will discuss the conclusions and limitations of the research.

Qualitative Data Analysis

This part was designed as a basic qualitative study with an interpretive approach in this research. The qualitative data comes from two sections. One is texting the from the official documents of the minors in program, the other section is the collected data through the survey open-ended questions. All the transcripts will be analyzed through standardized qualitative methods. The main purpose of the qualitative method is to

conclude the goal of the program, as well as to investigate the core factors that will affect the integrative learning of students as complementary materials for the refine of the framework. The main procedure of the qualitative data analyze will be coding. Coding is described (Charmaz, 2006) as "naming segments of data with a label that simultaneously categorizes, summarizes, and accounts for each piece of data". The text analysis will create focused codes that synthesized from the data and informed the development of categories. These categories then served as the basis for the thematic analysis, finally it will conclude the key themes or factors that will affect the integrative learning process.

The two open questions in pilot study are listed below:

Table 23

Open-Ended Questions in the Pilot Study

Item Description	Decision and Reason
What are 2-3 of your most positive takeaways from completing the minor?	Change Analysing Logic
What are 2-3 ways you think the minor could be improved?	Change Analysing Logic

The instrument was distributed by the Office of Undergraduate Education, the director preferred to collect more information on how to improve the program and student general education. The researcher asked two open-ended questions generally asked about how to improve their experiences in the program. Because the purpose of the pilot study was to test whether whole instrument worked, the researcher was totally open

to the data and entered the data without any assumptions, the method of coding data was more similar to grounded theory, it's an inductive way to develop a framework.

The data analysis process of qualitative research is inductive rather than deductive. The two questions in pilot study were very general, the researcher used inductive logic and found the answers were useful to get students' information on the integrative learning. After establishing the framework in the pilot study, the researcher revised the questions according to the framework of the instrument, try to use a deductive logic to make an inference to test whether the framework was mature enough.

Table 24

Open-Ended Questions in the Field Study

Item Dimension	Item Description
Integration within Minor	Please give one or two examples of times you made connections between courses in this minor.
Advancement through Integration	Please give one or two examples of times when you connected experiences in this minor with your experiences in the real world.
Integrative learning environment	Please give one or two examples how you interact with the people (faculty, students) in this minor.

The pilot study collected 2532 words; the three questions in field study collected 9067 words. In the second wave of field study, the qualitative questions changed back to "What are 2-3 of your most positive takeaways from completing the minor?" and "What are 2-3 ways you think the minor could be improved?" since the framework had been fixed in Spring 2011, the researcher plan to provide more information for the director to

improve the program. This round collected 3917 words in the second round of field study.

Coding

The inductive process will adopt coding logic in the pilot study since the questions are broad. The two questions in the pilot study did not bring any insights of integration to students, so the researcher will analyze the data from words to conceptive categories, then forms the frame without any assumptions. Although we do not large amount of text data in pilot study, it will still reveal the most important takeaways of the program from students' perspectives. From the process, we can test whether integration is a significant learning outcome of students, we can also figure out some important points that we may not conclude from the literature. The qualitative data in the pilot study help to construct a more completed framework for the instrument as complementary for the quantitative data. This process is like the coding process in grounded theory. "The strength of grounded coding derives from the concentrated, active involvement in the process. The researcher act upon data rather than passively read them" (Charmaz, 2006, p. 59). This coding approach needs researcher attention and passion.

The first step is to conduct initial coding while remaining open to explore data with any possibilities. Coding data without any assumption is a crucial point to construct new theory. The researcher code the data word-by word, line-by-line. Then is the theoretical coding, the literature review may arise theoretical foundation that can be integrated into the data, it implies the possible relationship between the categories you developed in the previous coding process.

For the field study, the researcher will use the deductive logic to do the qualitative data analysis since both framework and items were more mature than the pilot study and the open-ended questions were grounded in the structure of the instrument. This logic will scan the students' understanding of integration, help to validate whether the framework is effective to develop a consist assessment instrument.

This chapter introduced the process of establishing the initial framework, item development, data collection and plans to analyze the data. The three waves of data proposed and revised the framework for the assessment, it set up the foundation for the instrument development. The data collected that can be applied to item analysis in the following chapter, the item analysis result will give suggestions to each individual item.

Chapter 4 Results of Study

This chapter presented the results of analysis in multiple methods used to answer each of the research questions. The last chapter already introduced the initial framework of the instrument, based on that this chapter started with the qualitative research because the finding of the qualitative data collected in the pilot study and the wave one of field study was applied to decide the final items appeared in the wave two field study. Then, the research described the demographic information of the sample, combined the two waves of data, and report the results of the data analysis. After that, the findings are used to answer the research question.

Summary of Participants

The participants in this study came from the students enrolled in capstone courses of the Pathway Minor program in fall 2020, spring 2021, and spring 2022. The majority of the students were junior and senior students. This was expected since this research focused on the capstone course that occurs at the end of the program. Female students outnumbered male students which is not surprising considering the nature of many of the theme area aligned more closely with academic program with higher enrollment of women. However, since the capstone courses included the students from within a major, the researcher could not get an accurate count of the students enrolled in the minor. This meant I could not calculate response rates. The total numbers of participants by wave were: (1) 196 students in the spring 2021 of which 151 students completed it, (2) 106 participants in spring 2022 participated of which 77 students completed it. The overall completion rate for the survey was 75.5%. The data in Spring 2021 was gathered through

Qualtrics while the data in Spring 2022 was collected on QuestionPro. Although the software changed, the items remained the same. The number of respondents were different from the two waves; however, the demographic information of the two waves data aligns with each other whereby the majority of respondents were female, white, and juniors or seniors. The researcher did not participate in the data collection process of the Spring 2022 data. Since the survey was distributed by a shared link, the survey question "What is the Pathways minor you are completing this survey for?" helped to distinguish which class students enrolled. There were 75 of 77 respondents who answered the question during this wave. Accordingly, participants from 15 minors responded in spring 2021 and13 minors responded in spring 2022. One new minor was added in spring 2022 called Global Business Practices to Improve the Human Condition.

The researcher used descriptive statistics methods, independent samples t-test and CFA factor analysis loading to compare the field study data in spring 2021 and spring 2022. This analysis showed that the demographic information of two waves of data were consistent with each other. Since the means of each scale and factor loading did not have significant differences, the researchers combined two waves of data for the final analysis. However, since the enrollment number of the minor was unavailable, there may be a level of inaccuracy in participation rate of the assessment project.

Qualitative Data Analysis

The qualitative data is very important in this process of developing the instrument framework and items, this chapter started with qualitative method because it closely related to the final framework to collect the wave two data. The statistical method had

typically been complained and for use to determine the fit of a model to data for two primary reasons: the effects of sample size on statistical significance and the of inaccurate estimation of response in item due to insufficient sample sizes in certain ranges of attitude levels (Koch, 1983). The sample of this research was not very big from statistical perspective, it may affect the result of the model fit. While the quantitative analysis provided a general picture to establish and validate the instrument, qualitative analysis can also be used to validate the framework and improve the instrument.

Various qualitative methods have long been applied to the process of developing survey items (Castillo-Díaz & Padilla, 2013; Nassar-McMillan & Borders, 2002). The qualitative data in this study offered relevant insight into students' perception of integrative learning. For example, program proposals documents were essential materials and included data about the basic learning outcomes of the program. Similarly, the data collected in the open-ended questions were designed to find aspects in integrative learning that may have been obscured during the instrument development process.

Pilot study

The purpose of the pilot study was to examine item responses and insight to overall experiences with the integrated general education program, both goals leading to confirmation of the framework and improvements with the instrument. During this portion of the study, two open ended question were used to collect qualitative information:

1. What are 2-3 of your most positive takeaways from completing the minor?
2. What are 2-3 ways you think the minor could be improved?

The questions were generally asking students about positive and negative experiences with the program which meant that students could answer in ways that allowed them to express their insights in meaningful ways. These two questions did not include any presumptions and allowed the researcher to code the data without a specific standpoint to validate the established framework from the existing literature. In this wave of data analysis, the researcher used inductive coding. First, researcher concluded the excerpts for the raw data, then use the descriptive coding to conclude the meaning of the excerpts and categorized them into different categories, finally developed the themes. From the process of the analyzing qualitative data, the pilot study concluded several high frequencies excerpts through focused coding. Example excerpts are provided in Table 31.

Since the qualitative data in pilot study had no leading direction for students' responses, they could answer anything that they thought was appropriate. Mostly, it showed students' learning experiences, both good and bad, in the minor program.

From the table below, the research coded the open-ended text to common categories, then generalized the themes in the data." From the sample text, when talking about takeaways, students mentioned many essential points that correspond with factors of integrative learning prevalent in the literature. Further, responses presented students' knowledge learned in minor, and how they connect the knowledge to real issues, connect and apply knowledge to real world. When being asked about "What are 2-3 ways you think the minor could be improved", students also mentioned their dissatisfaction on course construct. Through data coding process, most of the takeaways and dissatisfaction were related to the factors of integration we established in literature review, as the table

showed. The theme of the qualitative data of pilot study were “Integration within Minor”, “Integration beyond Minor” which were quite align with the dimensions in the framework. They were showing the framework the researcher established was effective for the instrument to assess the integrative learning in this program.

Table 25

Examples of Coding from Excerpts to Themes

Excerpts	Categories	Themes
Excerpt 1 The capstone has a difficulty jump for minor students who only has seldom course experiences before. The capstone course feels like a big jump for students in the minor to make, since we previously have only taken one course, so I think it would be helpful to have introduce an intermediate class or make the capstone course separate for minors and majors. **Excerpt 2** Minor should provide more foundation knowledge because students from other major have difficulty in understanding the course. I think that the minor should focus on the basics in the intro. Because I am a biology major, I found it hard to write a philosophical paper without structuring it the same as an English paper. Some of the topics also went above my head. I felt uncomfortable speaking in class at first because I felt like anything I said wasn't up to par with the rest of the class. I am much more comfortable now and don't mind speaking up, even if I am scared, but that was after the first minor class. I spent a lot of time in office hours just trying to understand the readings. **Excerpt 3** One class does not fit with other classes. **Excerpt 4** Direction of the minor is not clear The Eco Ag class does not really fit with the other classes in the minor. It didn't emphasize the community partners as much and the direction of the course itself wasn't very clear to me. **Excerpt 5** The content in the courses are overlap Also, I think this course has a lot of overlap with intro to disability studies, so I would consider taking out some information or going over most of the introductory disability vocabulary and information within the first couple weeks of class.	The problem of course construct	Integration within Minor

Excerpt 1 The knowledge learned in the class can make student know more issues in the real society I find the class very fascinating and enjoy the material but a reason why I loved introduction to disabilities studies is that it allows you to understand both sides to many of the issues disabled people face and the relationship they have with nondisabled people in society.	Connect knowledge to the real-world issues	Integration beyond Minor
Except 1 Interactive course provided information of different ways to solve problem in the real world. I also really enjoyed the material of the courses I took within this minor. It was always interactive and provided me with so much information about how we can use agriculture in different ways to fix food insecurity in the world. Except 2 Hear knowledge from people with different background Hearing the perspectives of people who are actually disabled and not just professionals learning about assistive technology After taking courses for the Disabilities Studies minor, I see people who are disabled with a different mindset than I did before.	Different perspective to see a topic Different ways to deal with a problem	Integration beyond Minor
Excerpt 1 Make connections with others different than me Being able to relate to others different than me.	Connect others different than me.	
Excerpt 2 Hear important knowledge from people with different backgrounds to know different perspective on a certain topic. I was able to hear from many important speakers that have different backgrounds. It showed me that there are many ways to combat Global food security and health.	Learn knowledge from people with different backgrounds	
Except 1 Apply knowledge learned in minor to future career Being able to use this knowledge I learned in my future career.	Apply knowledge to future career.	Integration beyond Minor

Except 1 Collaborating with my peers hearing the perspectives of people who are actually disabled and not just professionals learning about assistive technology	Collaborate with peers	Integrative Learning Environment

Field Study

In the field test, the researcher used deductive logic to revise the questions according to the framework of the instrument. The researcher developed the predetermined codes based on the framework of the instrument. Three questions were concluded from the scales of the framework:

1. Please give one or two examples of times you made connections between courses in this minor. (Integration within Minor)
2. Please give one or two examples of times when you connected experiences in this minor with your experiences in the real world. (Advancement through Integration)
3. Please give one or two examples how you interact with the people (faculty, students) in this minor. (Integrative Learning Environment)

The questions were concluded from the framework, the first question was asking students' experiences of the integration within minor, the second question was asking the experiences of advancement through integration. The final question was asking the environment climate in the program. In this wave of data collection, students answered the qualitative questions from these three dimensions.

We set these three scales as the main themes, then each theme was divided into several subthemes.

The researcher used the subthemes to code the qualitative data transcript. After reading word-by word, line-by-line, many important points that ignored in the original instrument came to the surface.

Table 26

Theme and Subthemes Used in Coding Process

Themes	Subthemes
1. Integration within Minor	1.1 Students' perceptions of course construct 1.2 Students' behaviors of knowledge integration across the program
2. Advancement through Integration	2.1 Student behaviors of integration with real world issues. 2.2 Student behaviors of integration with other experiences. 2.3 Student behaviors of integration with future.
3. Integrative Learning Environment	3.1 Faculty student interaction. 3.2 Peer interaction. 3.3 Social engagement.

Students' perceptions of course construct **(subtheme 1.1)** was discussed in the open-ended question. For this question, students shared their understanding the most important factors in course construct in a theme-based program. Some of students talked about their experiences generally.

For example, one student wrote, "All the R courses were extremely interconnected and built off of one another."

Many students emphasized the vertically connections between courses in the program. One student's comment is very representative of others, "The very first course informs you on the general knowledge of greenspaces and ecological principles. Then in this course we were able to further explore these and learn how to best apply them"

In the instrument we asked about the commonalities and differences between the courses, but the instrument did not pay attention to the vertical construct within the minor courses.

Students' behaviors of knowledge integration across the program **(subtheme 1.2)** were defined as important integration dimension.

For this sub theme, many students discussed the course content overlap, some students talked about their experiences to apply knowledge across different fields, "Many of the themes discussed in classes are similar, especially the concepts for critical thinking and systems thinking"

One student mentioned how she used the knowledge in different fields, "In my technology class and my design class I have been able to use the knowledge of both together."

Student behaviors of integration with real world issues (subtheme 2.1) shared students' connection and application of knowledge to the problems related to the minor theme. Students shared their real project in their learning process:

> "I found that class to be very informative and it related to a lot of real-world issues. For example, I learned a lot about the harmful effects of meat production in the US and have lowered my consumption of meat due to it. I have also been able to recognize different parts of the Stroubles Creek watershed around campus."

Student also shared how he/she identified issues with knowledge they learned from minor, “I work with a student who has down syndrome and I have identified places on campus that are not accessible to her.”

Student behaviors of integration with other experiences (subtheme 2.2) showed connections that students make between minor and their other information.

One student told the researcher how he/she applied knowledge from minor to major, “In writing my final paper for the capstone course, I used concepts I learned in the minor intro course as well as in my major.”

One student showed how he/she made connections between the knowledge he learned and real business,

> “I have made connections just in different business ideas and how to grow my own landscaping business and I have also learned how to think and create business models which help me think through my new business ideas.”

The theme “Integrative Learning Environment” which was dimension four, was assessed whether the program environment was supportive for students’ integrative learning. This theme includes three sub-themes.

Faculty student interaction (subtheme 3.1) This subtheme considered students interactions with instructors inside and outside of the classroom.

One student said the faculty helped he/her in the study, “Faculty was always willing to meet at a time that was convenient to me and always made enough time to ensure I grasped a full understanding of the material.”

Peer interaction (subtheme 3.2) This one considers different formats of peer interactions between students such as group projects, class discussions.

Here is a student shared the peer interaction in the learning, “In an academic realm, I was able to collaborate on group projects with my peers that we are Engineers, College of Business, and CS/CMDA majors.”

Social engagement(subtheme 3.3) From 1984, Astin(Astin, 1984, 1993) continuously proposed involvement theory to emphasized the importance of social aspects of students’ learning experiences.

Student shared the experience of establishing relationship outside of the classroom, “In a non-academic realm, I was able to interact with my peers outside of the classroom and establish interpersonal relationships.”

One other student said, “I have made friends in this minor who I interact with outside of class.” (*Establish friendship, subtheme 3.3*)

Results of Qualitative Data

The first, and central, finding from the qualitative data analysis is that the framework of the instrument is effective to assess the integrative learning in minor program. Data from both pilot study and field study point to the instrument being cohesive and understandable to the respondent. The data of pilot study was analyzed by using initial coding followed by focused coding. Combined, these two coding methods were used to validate the framework that was established from the literature review and the minor proposal documents. Then the field study used proposed frameworks as

theoretical key words to analyze the data. These processes validate the framework of the instrument and were used to make minor revisions to it which are described next.

From the qualitative analysis process, the data suggested that it was necessary to add new items based on the existed framework to improve it. These were reflection, innovation, and course construction.

Reflection

The instrument had an item on the reflective assignment, but not attention to careful considerations which lead the transformations. Here is an example from a student:

> "Coming from a STEM perspective, this minor made me rethink a lot of the things engineers design without consulting their actual intended audience. All of these courses left me with my jaw hanging open at the new information I learned and new perspective that was shared with me day after day. My friends and family can attest to the numerous conversations I've brought up related to since starting this minor. Most of the videos and TedTalks from these classes have been passed on to them. My friend and two of my stepsisters work in special education classrooms and we will often talk about the things I've learned in this class and even more varying opinions on how to best help and teach people with disabilities while also respecting them as individuals who can decide what they want too. Post-graduation, I knew I wanted to combine my love of programming and robotics with a desire to help people, but throughout these classes, I've learned there's good and bad ways to 'help' and I would not have known that if not for these classes, so thank you for having this minor!"

The answer showed the whole process how he or she reflect the knowledge learned from minor, this process included his own connection between, discussion with others, full reflexivity between the minor topic and his own experiences, then changing of his mind and future career. The data suggested that we can ask more on students' experiences of reconfiguring their thought systems, establishing the accomplishment of learning goals, organizing learning process, and meaning making. The items should focus more on students' cognitive practices of re-examining the learning process through the cognitive, intrapersonal, and interpersonal domains.

The instrument could add some questions about students' reflection by themselves, at the same time, consider how instructors ask questions and create time to provoke students to find connections, value independent and critical thinking develop consecutive understandings in separate disciplines of knowledge and skill.

Innovation:

In the qualitative data, students mentioned the word "new" many times, including new ideas and new solutions.

For example, several students mentioned the takeaways from minor were innovative ideas about solutions in real life, "In both the Create! Course and the Startup: I completed semester long projects that required teamwork and had us take an idea and help develop solutions throughout the semester."

This student said, "I have also learned how to think and create business models which help me think through my new business ideas."

The instrument has some items on how to connect and apply knowledge in new context, they are the foundation of innovation. For example, one of item states “I have used the knowledge gained from this minor to solve problems in my everyday life.” This is similar to the innovation capabilities. Integrative learning and innovation are theoretically and practically connected, Mayhew, Selznick and colleagues (Selznick et al., 2022) said innovation is a set of competences and abilities linked with involved in the process of producing and applying new ideas contextually. How to transfer and apply knowledge to new context is also crucial to innovation. From this perspective, innovation is a high cognitive learning outcome of integrative learning. We need to add it to the instrument. So the item BEYMIN6 was added to the instrument.

Course Construction:

Although in the instrument, the researcher developed several items on different dimension of the course construction, for example: “the number of classes I completed for this minor is sufficient to meet the goals of the minor”. But many students mentioned the problems on course construction, content overlap, and content consistency. For example, “*The content for Abnormal Psychology and Developmental Psychology overlapped at times.*” For course construction dimension, students mentioned the elective courses were not enough for them to choose. Students also proposed that the foundation of knowledge was hard for students to understand without knowledge backgrounds. In addition to this, students also pointed out the content difficulty gap between the foundation and elective classes and the capstone courses. For this part of the items, more

items need to be added to the instrument to evaluate the whole construction of the courses especially the difficulty gap within the courses.

After analyzing the qualitative data, the researcher decided to add three questions to the instrument of wave two. One item was on the continuousness of the course, which was THEME 5, one was on innovation which was BEYMIN 6, and one more question on students' own reflection called KNOWLEDGE5.

Results of Quantitative Study

Following the steps outlined in Chapter 3, I narrowed the original sample of respondents from 302 to 228. 196 students participated in the Spring 2021 survey, 151 students completed it, 106 participants contributed answers to the Spring 2022 survey, 77 people completed it. During the data cleaning process, for spring 2021 data, the researcher selected the finished variable as "finished", selected the progress variable ad "100" in Qualtrics which means the percentage of instrument that the respondents have completed as 100%, then the researcher checked duplicates and eliminated them by SPSS syntax. The data in Spring 2022 was collected on a platform called QuestionPro since the university stopped cooperation with Qualtrics. The variables "Finished", "Status" and "Progress" particularly designed by Qualtircs to clean data have been removed. The data was collected and managed by another graduate assistant. After closing the survey, the graduate assistant shared the raw data with the researcher. Through raw data, the researcher can only delete the cases by detecting missing data and duplicates. The SPSS was still used in this round of data cleaning

Table 27 reports the demographic characteristics of the sample. The majority of the sample was female (71%) versus male (28%), White or Caucasian (not Hispanic) was taking 72.8% of the sample while Asian or Pacific Islander was taking 28% and 20% of the participants were more than one of the races. Since our sample were all coming from the capstone courses, most students were juniors (18.0%) and seniors (73.2%).

First, the researcher calculated the reliability of the instrument and each scale, compared the IRT models indices to find the model most appropriate to the data, then used the model to do the item-level analysis. The processes showed the quality of both the instrument and the items from a statistical perspective. After that, the qualitative data in pilot study and first wave of field study is applied to revise the instrument. The qualitative analysis described the coding process, lists the excepts with high frequencies, then concluded with suggestions to improve the instrument.

Table 27

Demographic Information of Participants

Question	Response	*n(%)*
Sex	Male	63(27.6%)
	Female	163(71.5%)
	Prefer not to answer	2(0.9%)
	Total	228(100%)
Race/Ethnicity	Asian or Pacific Islander	28(12.3%)
	Hispanic	9(3.9%)
	Black or African American	12(5.3%)
	White or Caucasian (not Hispanic)	166(72.8%)
	More than one of these	20(8.8%)
	Another race/ethnicity not provided here	2(0.9%)
	Prefer not to answer	1(0.4%)
	Total	228(100%)
Class Standing	Freshmen	3(1.3%)
	Sophomore	14(6.1%)

Junior	41(18.0%)
Senior	167(73.2%)
Missing	1(0.4%)
Total	228(100%)

Reliability within Classical Test Theory

The first wave and the second wave of field study were combined by SPSS syntax. The items in the final version of the instrument are listed in Table 28. Data was combined using responses in both spring 2021 (wave 1) and spring 2022 (wave 2). All items in both of the first wave and second wave were analyzed. Several items THEME4, KNOLEDGE5 and BEYMIN6 developed by the qualitative data result, they only appeared in Wave 2 of the field study will not be analyzed in the quantitative method. The numbers of the item were adjusted based on the wave one of data, the same content was named the same. Because the researcher deleted some items, so the numbers of the item were not continuous. There were only 77 participants, quite different from other common items. To improve data accuracy in the quantitative analysis, the researcher decided to keep these three items separate and introduce a development process in the qualitative analysis.

Table 28

Construct of the Instrument

Scales	Item	Description
Integration within Minor (IWM)	THEME1	Course Theme
	THEME2	Course Quantity
	THEME3	Course Construct
	THEME4	Course Continuousness
	DISGR1	Disciplinary Grounding
	DISGR2	Disciplinary Grounding

	DISGR3	Disciplinary Grounding
	DISGR4	Disciplinary Grounding
	KNOWLEDGE1	Knowledge Application
	KNOWLEDGE2	Knowledge Application
	KNOWLEDGE3	Knowledge Integration
	KNOWLEDGE4	Knowledge Integration
	KNOWLEDGE5	Knowledge Reflection
Integration beyond Minor (IBM)	BEYMIN1	Integration to Real World
	BEYMIN2	Integration to Real Problem
	BEYMIN3	Connections to Previous Knowledge
	BEYMIN5	Connection to Future Career Plan
	BEYMIN6	Connection to Innovation
Integrative Learning Environment (ILE)	LE1	Environment Support
	LE2	Instructor Support
	LE3	Faculty Interaction
	LE4	Reflective Environment
	LE5	Reflective Environment
	LE6	Peer Interaction
	LE8	Peer Interaction
	LE9	Peer Interaction

All the quantitative result were calculated using 23 items excluding THEME4, KNOWLEDGE5 and BEYMIN 6.

One of the most valuable conceptions in CTT is reliability. The term focuses on accuracy in measurement, and it is described as consistency and steadiness of output scores over recurring measurements (Brennan, 2001).

The framework of instrument indicated that the items in the instrument were not testing one subject, especially where Integrative Learning Environment was measuring different things from other two scales that mainly concentrated on an integrative learning process. Next, the researcher tested the unidimensionnality of each dimension in the framework. After confirmation of the construct, we used jMetrik software to calculate reliability in each of four components of integrative learning. This approach was taken

because we aim to produce scores of each subscale through the instrument for rather than a single measure of integrative learning.

Assessment of Unidimensionality. Testing the model fit is needed to be built on the different scales generated by the instrument framework. To test the assumption of the unidimensionality, each scale (for a total of 3 scales) was assessed by confirmatory factor analysis (CFA).

The Table 30 is presenting the CFA result, both model fit and principal component. The test was done in both SPSS and R. SPSS presented the result of PCA eigenvalue, R shows the summary of the indicators. The minimum eigenvalue of the first principal component is the scale of Integrative Learning Environment which is 31.137% of the total scale variance and a maximum of 67.102% of the total scale variance for Integration Beyond Minor. The scale of Integrative Learning Environment showed more than one eigenvalue greater than one. The Scree plots for these analyses showed the proportion of total variance accounted for by the second eigenvalues. In Scale IWM and Scale IBM there was only one eigenvalue in each of them. Special attention was paid to the second eigenvalues of Scale ILE which was 24.251%. This is not a small proportion, and the existence of additional components may lead a lower reliability for this scale.

The results of the confirmatory factor analysis show a reasonable fit for a one factor model for the two scales Integration within Minor and Integration beyond Minor but not very good for the third scale Integrative Learning Environment. The higher values of GFI which is scaled from 0 to 1 suggesting better model fit. The cut-off value of GFI was set to 0.9 (Joreskog & Sorbom, 1984). The maximum of GFI value were between

0.990, indicating the fit is acceptable. The first two of RMR were all less than 0.05, indicating a good fit(Byrne, 2013). Although the RMSEA's for first scales was over 0.1, based on the other evidence of PCA, GFI and RMR presented, these were still reasonably small enough to consider for the first two scales. The one factor model fit the data.

The assessment of unidimensionality for the Scale GC was not acceptable. Analysis in R resulted in error messages indicating fit measures were not available because the model did not converge. This is an indicator that at least some items have variances that is very dissimilar to the others and that the inner construct of scale was not uniform. The researcher decided to use Exploratory Factor Analysis (EFA) to explore the construct to test the factor loading. The factor loadings of LE3, LE4, LE5, LE6, LE8 and LE9 for factor one was greater than 0.50. The factor loadings of LE1 and LE2 for the factor two demonstrated two items with loadings were greater than 0.90 which are very good loadings. The result was presented in Table 31.

Table 29

CFA Fit Statistics for the Initial Framework

Scale				CFA Fit Statistics One Factor Model		
	No Eigen. over 1	% Var. First Eigen.	% Var. Second Eigen.*	GFI	RMR	RMSEA
IWM	1	62.613		0.892	0.029	0.104
IBM	1	67.102		0.990	0.017	0.079
ILE	2	31.137	24.251			

The EFA conducted in this research showed the last scale can be divided into two scales. The items of LE1 and LE2 have been divided into the scale named General Climate (GC), the items LE3, LE4, LE5, LE6, LE8 and LE9 have been categorized as

factor two Environmental Support(ES). Then the researcher did the unidimensionality test again, the results indicated the four-scale construct was better than the three-scale one.

Table 30

Rotated Component Matrix[a]

	Component	
	1	2
LE1		.974
LE2		.982
LE3	.605	
LE5	.569	
LE6	.741	
LE4	.727	
LE8	.544	
LE9	.649	

Extraction Method: Principal Component Analysis.
Rotation Method: Quartimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

Table 31

CFA Fit Statistics for the Final Framework

Scale	PCA Results			CFA Fit Statistics One Factor Model		
	No Eigen.over 1	% Var. first Eigen.	% Var. second Eigen.	GFI	RMR	RMSEA
IWM	1	62.613		0.892	0.029	0.104
IBM	1	67.102		0.990	0.017	0.079
GC	1	96.478		1	0.000	N/A
ES	1	41.449		0.966	0.040	0.085

All the indicators showed very good fit when dividing the Integrative Learning Environment (ILE) scale into two and the assumption of unidimensionality had been satisfied. Interpretation in these cases is always a combination of mathematics and including a judgement of the population and of the range of contexts of the phenomena (Kane, 2010). In this study, I made own decision to maintain all the items into the analysis.

Reliability of Each Scale. Next, the researcher examined the reliability of each scale. Table 33 shows the results of the reliability analysis of each scale.

From the table we can conclude the radiality of each scale were acceptable. The scale one was the largest which was over 0.9, the Scale GC was also over 0.9, this may be resulted from only two items in the scale. The Scale IBM was over 0.8 which was also very good. The reliability of Scale ES was over 0.7, not very high but still passed the statistical standards.

Table 32

Reliability Analysis

Scale	Method	Guttman's L2	Coefficient Alpha	Feldt-Gilmer	Feldt-Brennan	Raju's Beta
IWM	Estimate	0.9489	0.9484	0.9487	0.9487	0.9484
	95% Conf.	(0.9385, 0.9582)	(0.9379, 0.9578)	(0.9383, 0.9581)	(0.9383, 0.9581)	(0.9379, 0.9578)
IBM	Estimate	0.8459	0.8430	0.8478	0.8477	0.8430
	95% Conf.	(0.8104, 0.8762)	(0.8068, 0.8738)	(0.8128，0.8777)	(0.8126, 0.8776)	(0.8068, 0.8738)
GC	Estimate	0.9639	0.9639	N/A	0.9639	0.9639

	95% Conf.	(0.9531, 0.9722)	(0.9531, 0.9722)	N/A	(0.9531, 0.9722)	(0.9531, 0.9722)
ES	Estimate	0.7201	0.7162	0.7196	0.7175	0.7162
	95% Conf.	(0.6599, 0.7729)	(0.6550, 0.7697)	(0.6593, 0.7725)	(0.6567, 0.7707)	(0.6550, 0.7697)

IRT Model Fit

Comparisons Across Models. Next, the researcher explored model fit. The polytomous IRT models are for items that have more than two possible scores, the most common examples are Likert-type items. In the measure developed in this study, there were two types of rating scale items, one was 6-point Likert scale, the other one was a 4-point Likert scale. The scale one, two, three used a 6-point Likert scale. Scale ES used a 4-point Likert scale. The Rating Scale Model requires the Likert items instrument which elicit the same rating structure across items, the Scale ES was 4-point scale which was different from the previous three scales.

As a result, in this research, four IRT models were ran and compared for the goodness-of-fit statistics, they were Partial Credit Model (PCM), Graded Response Model (GRM), General Partial Credit Model (GPCM) and Nominal Response Model (NRM). To compare these models, the following model-fit indices were applied: Akaike Information Criterion which we known as AIC (Akaike, 1974), Bayesian Information Criterion that is called BIC (Schwarz, 1978). Other research, for example New Developments in Quantitative Psychology (Millsap et al., 2016) proposed that the BIC was more preferred in model fitting. The generalized S-$\chi 2$ in the process was also used to measure item fit. The values were compared across the four models for the same datasets

combined from the wave one and wave two of field study. Comparisons were made on the same items across the different item subsets when the same model was applied. Decisions were made by the consideration of all three indices. These parameters were calculated using IRTPRO 4.2 Version. Results of these tests are shown in Table 18.

Table 33

Comparisons Across the IRT Models

Model		-2ln L	AIC	BIC
PCM	Scale IWM	4280.27	4390.27	4579.12
	Scale IBM	1953.53	1993.53	2062.21
	Scale GC	1314.10	1334.10	1368.44
	Scale ES	3169.87	3205.87	3267.68
GRM	Scale IWM	4050.32	4182.32	4408.94
	Scale IBM	1910.46	1958.46	2040.87
	Scale GC	1142.77	1166.77	1207.98
	Scale ES	3139.51	3187.51	3269.92
GPCM	Scale IWM	4248.18	4380.18	4606.80
	Scale IBM	1910.10	1958.10	2040.51
	Scale GC	1165.15	1189.15	1230.36
	Scale ES	3143.55	3191.55	3273.95
NRM	Scale IWM	4296.52	4516.52	4894.23
	Scale IBM	1991.80	2071.80	2209.14
	Scale GC	1163.43	1203.43	1272.10
	Scale ES	3134.27	3206.27	3329.89

Table 33 shows results from the likelihood ratio, the AIC and BIC values of the Scale IWM, Two and Four. In most of the indices, GRM were smallest, revealing fitted significantly better than any other models. For Scale ES, the smallest likelihood and AIC value was still GRM while smallest BIC was PCM. Graded Response Model was overall the best fitting model for all scales. Considering the uniform of the models for three other scales, it was chosen to analyze the instrument in the following analysis.

Item Analysis. After the analysis at the instrument level, then the research comes to the item level analysis. The item analysis adopted polytomous IRT models. Three different methods of item analysis were utilized in this research: (1) traditional summarized statistics; (2) parameters estimation, and (3) IRT plot.

Since the answers to some items did not have choice of 2, all the software no matter IRTPRO, Jmetrik or R distribute the intervals automatically from 6 categories to 5 categories. To solve this problem, the researcher made a fake case which only includes the choice of 2 for all the variables.

In table 35, the summarized statistics for the item in the Scale IWM is presented. Since this instrument was regarded as a norm-referenced score instrument, the standards for the measurement development desire items with high item-total correlations, high standard deviations and items without extreme difficulty or low discrimination. From the table 35, each of the items have five category boundaries which formed six response categories. The average values of responses of each item were not far away from each other. The biggest was THEME 2, which is 4.375, the smallest was DISGR4, which is 4.098. The Item-Total correlation suggests if any of the items do not have responses that did not align with those for other items. In this scale, all the values were greater than 0.6 indicating the coefficient were excellent. The indices suggested the scale discriminate well, and it is a nice construct of the measurement to find out what participants' experiences. When deleting the item, the Coefficient α all decreased, this means all the items were important to the scale. The category boundaries, discrimination parameter estimates for the items and abilities for the participants were obtained from IRTPRO

which was designed to estimate parameters for a variety of polychotomous item response models.

Table 34

The Summarized Statistics of Raw Data of Scale IWM

	Response		With Item Deleted	
			Item-Total	Coefficient α
Item	Average	Std. Dev.	Correlation	
THEME1	4.295	0.838	0.8030	0.9378
THEME2	4.375	0.771	0.6684	0.9427
THEME3	4.321	0.844	0.7661	0.9392
DISGR1	4.138	1.017	0.7317	0.9411
DISGR2	4.250	0.858	0.7808	0.9386
DISGR3	4.228	0.807	0.8108	0.9377
DISGR4	4.098	0.873	0.7204	0.9409
KNOWLEDGE1	4.304	0.877	0.7582	0.9394
KNOWLEDGE2	4.254	0.880	0.7804	0.9386
KNOWLEDGE3	4.366	0.842	0.7718	0.9390
KNOWLEDGE4	4.121	1.015	0.7613	0.9397

Table 34 presents parameter estimation of the Scale IWM calculated by grade response model. Data in Table 36 show the estimated parameters of the items in the Scale IWM. The item discrimination values ranged from 2.14 to 3.50 and were randomly assigned to the items. Each of the items had five category boundaries which formed six response categories. The biggest α was 3.50 belonged to the item DISGR 3, the smallest α was 2.14 belonged to the item THEME2.

The parameters b_5 of the item THEME2, THEME3 and KNOWLEDGE1 were 0.02, -0.01 and -0.01, close to 0. It meant the category 6 in this item has small effect.

Table 35

Graded Model Item Parameter Estimates for Scale IWM, logit: a(θ - b) *(Back to TOC)*

Item	a	s.e.	b1	s.e.	b2	s.e.	b3	s.e.	b4	s.e.	b5	s.e.
THEME1	3.21	0.41	-3.93	0.73	-2.64	0.34	-2.10	0.22	-1.20	0.12	0.07	0.09
THEME2	2.14	0.28	-4.16	0.75	-3.01	0.41	-2.63	0.33	-1.79	0.20	0.02	0.10
THEME3	2.89	0.37	-3.96	0.73	-2.76	0.37	-2.25	0.25	-1.25	0.13	-0.01	0.09
DISGR1	2.50	0.31	-2.72	0.35	-2.29	0.26	-1.81	0.19	-1.13	0.13	0.25	0.10
DISGR2	3.02	0.38	-3.97	0.73	-3.08	0.45	-1.76	0.18	-1.25	0.13	0.17	0.09
DISGR3	3.50	0.45	-3.92	0.72	-3.01	0.43	-2.18	0.24	-1.12	0.11	0.22	0.09
DISGR4	2.58	0.31	-4.05	0.74	-3.18	0.46	-1.97	0.21	-0.98	0.12	0.40	0.10
KNOWLEDGE1	2.67	0.33	-3.98	0.72	-2.75	0.35	-2.19	0.24	-1.14	0.13	-0.01	0.09
KNOWLEDGE2	3.07	0.39	-3.95	0.73	-2.68	0.35	-1.97	0.20	-1.09	0.12	0.07	0.09
KNOWLEDGE3	2.85	0.37	-3.97	0.73	-2.49	0.30	-2.31	0.26	-1.33	0.14	-0.11	0.09
KNOWLEDGE4	2.95	0.36	-3.21	0.49	-2.60	0.33	-1.67	0.17	-0.82	0.10	0.14	0.09

Table 36

S-X^2 Item Level Diagnostic Statistics for Scale IWM

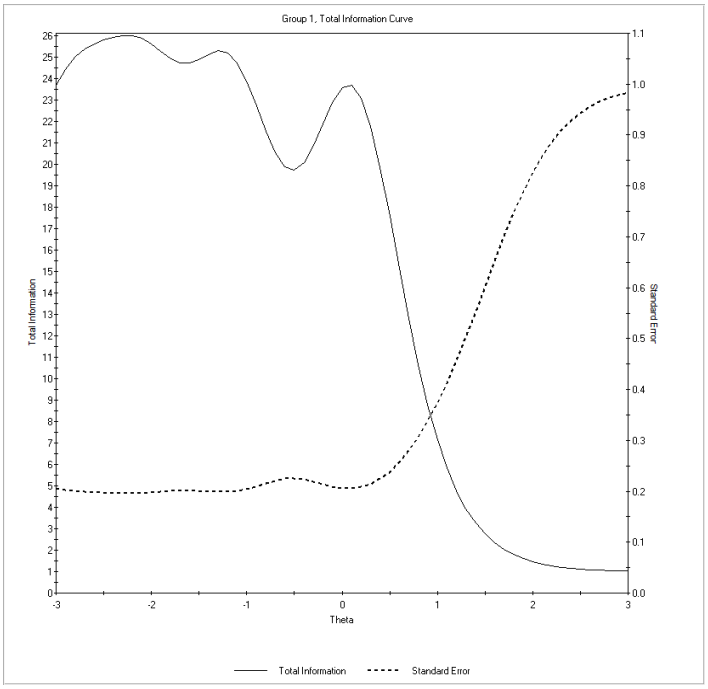
Item	X2	d.f.	Probability
THEME1	21.58	21	0.4254
THEME2	23.17	21	0.3339
THEME3	33.64	21	0.0395
DISGR1	30.84	29	0.3748
DISGR2	28.71	26	0.3236
DISGR3	18.85	21	0.5956
DISGR4	33.00	28	0.2352
KNOWLEDGE1	27.03	26	0.4093
KNOWLEDGE2	23.49	24	0.4927
KNOWLEDGE3	39.35	24	0.0250
KNOWLEDGE4	54.88	27	0.0012

Item level diagnoses (Table 36) provide chi-square tests of the fit of the GRM to each item in the scale. The chi-square statistics calculated from the observed and expected frequencies of responses in the item categories. From Table 37, of the 11 items in the Scale IWM after the item analysis, only one item (KNOWLEDGE 4) was found to have a lack of fit to the model because the probability is smaller than 0.01. Overall, the items in the Scale IWM fit the graded response model.

Figure 4 shows a chart of the total information curve of the Scale IWM.

Figure 4

Total Information Curve of Scale IWM



This total information curve shows item pool information functions, all items in this scale are 6 points Likert scale. The curve itself had three peaks, it ranged constant little up and down from -3 to 0. The information functions reached their maximum at approximately θ=-2.2, the second peak appears at approximately θ=-1.4. After the second peak point, the curve started to drop, but it rose again when θ=-1.0, after it achieved another peak at approximately θ=0, which is lower than the first one, it dropped off rapidly from 0 to 3. The greatest measurement precision was falling at around -2.2.

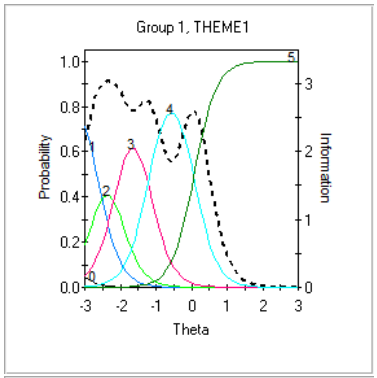
The item parameters determine the shape and location of the category response curves and operating characteristics curves. The narrower and more peaked the category response curve indicates that the response categories discriminated among trait levels well. The charts were combination of trace lines and item information curves for the items in the Scale IWM.

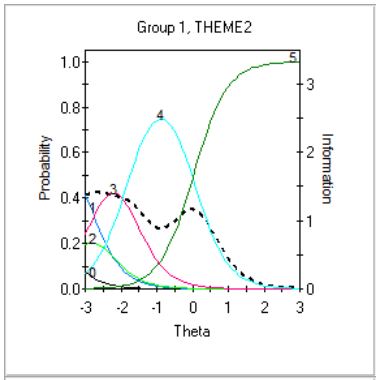
From the information curve, we can see the item THEME2 provided less information than others, the information curve was constantly lower than 0.4. The DISGR1 and DISGR4 started smaller than 0.5, then went up and down. For the trace lines of category response curves, the THEME2 were flatter than other items, which meant this item differentiated among trait levels worse than other items. Instead of scoring 1 to 6 in the raw data, the lines scored them from 0 to 5. The 0 line which means "strongly disagree" of all items were very low and narrow except DISGR1, the 5 which means "strongly agree" started in this scale, students showed extremely positive attitude upon the integration within minor, the 5 lines of all the items almost the same.

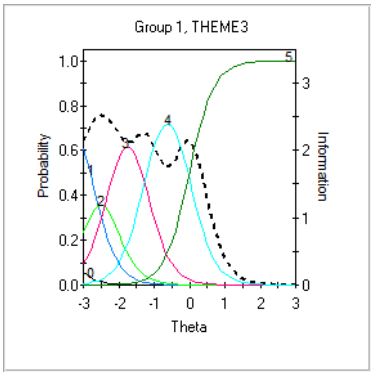
According to the information and analyzing results above, the KNOWLEDGE4 was not as good as other items but are not very bad which need to be deleted. But the parameters and the information curve which was lower than other items showed the item THEME 2 was not good as others but it was still normal. Overall, the Scale IWM fit the grade response model well.

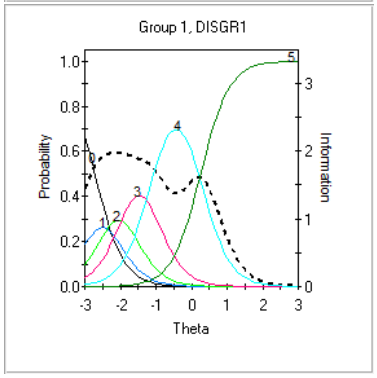
Figure 5

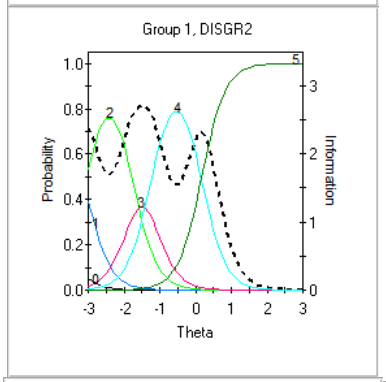
Combined Information Curve and Trait Links for Scale IWM

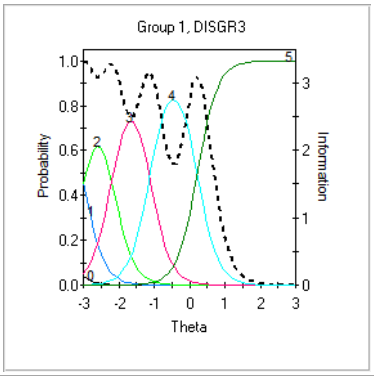

Group 1, THEME1
Probability
Information
Theta

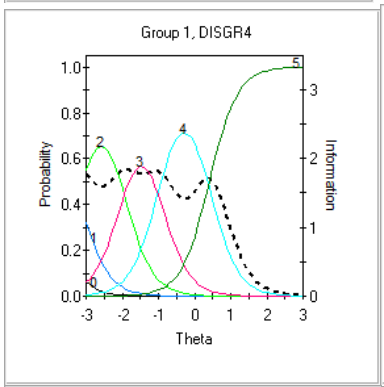

Group 1, THEME2
Probability
Information
Theta

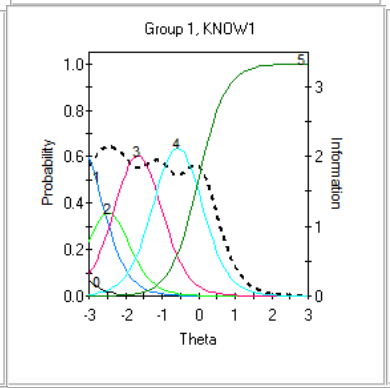

Group 1, THEME3
Probability
Information
Theta

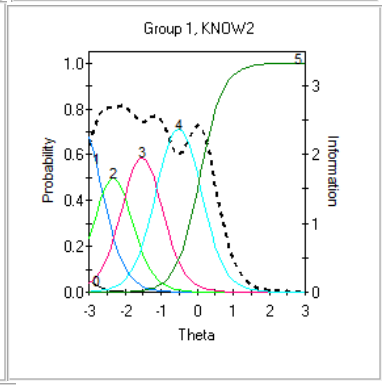

Group 1, DISGR1
Probability
Information
Theta

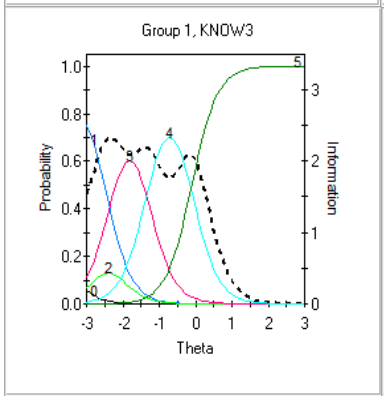

Group 1, DISGR2
Probability
Information
Theta


Group 1, DISGR3
Probability
Information
Theta


Group 1, DISGR4
Probability
Information
Theta


Group 1, KNOW1
Probability
Information
Theta


Group 1, KNOW2
Probability
Information
Theta


Group 1, KNOW3
Probability
Information
Theta

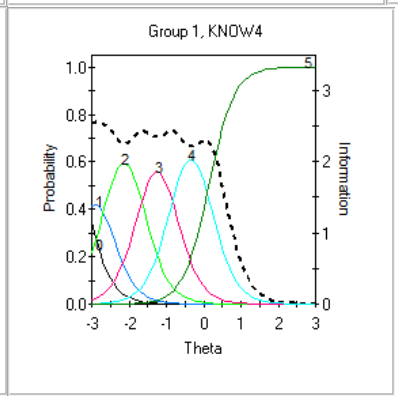

Group 1, KNOW4
Probability
Information
Theta

Table 37 shows the summarized statistics of the Scale IBM. The Scale IBM had four items. The biggest value of average was BEYMIN3 which was 4.216, the smallest was BEYMIN2 which was 3.687, and the standard deviation of BEYMIN2 was the largest which was over 1. The Item-Total correlation were not as high as the scale one but were still higher than 0.6 which were good. The indices suggested the Scale IBM also has a very good discrimination.

Table 37

The Summarized Statistics of Scale IBM

	Response		With Item Deleted	
Item	Average	Std. Dev.	Item-Total Correlation	Coefficient α
BEYMIN1	4.216	0.893	0.6878	0.7996
BEYMIN2	3.687	1.099	0.7059	0.7936
BEYMIN3	4.370	0.828	0.6461	0.8185
BEYMIN5	4.088	1.027	0.7011	0.7924

Table 38

Graded Model Item Parameter Estimates for Scale IBM, logit: a(θ - b) *(Back to TOC)*

Item	a	s.e.	b1	s.e.	b2	s.e.	b3	s.e.	b4	s.e.	b5	s.e.
BEYMIN1	2.59	0.39	-3.61	0.56	-3.11	0.40	-1.93	0.20	-1.13	0.13	0.14	0.10
BEYMIN2	2.62	0.37	-3.17	0.42	-2.08	0.21	-1.26	0.14	-0.37	0.10	0.78	0.12
BEYMIN3	2.33	0.37	-3.68	0.58	-3.13	0.42	-2.33	0.26	-1.42	0.15	-0.09	0.10
BEYMIN5	2.38	0.35	-3.02	0.39	-2.41	0.26	-1.94	0.20	-0.88	0.12	0.25	0.11

The *a*-values were quite close, the smallest was 2.38, the biggest was 2.62, not a huge range. The other parameters were reasonable and acceptable. From the parameter estimation, the parameters of the items in the Scale IBM met the requirements of good fit to the model.

Table 39

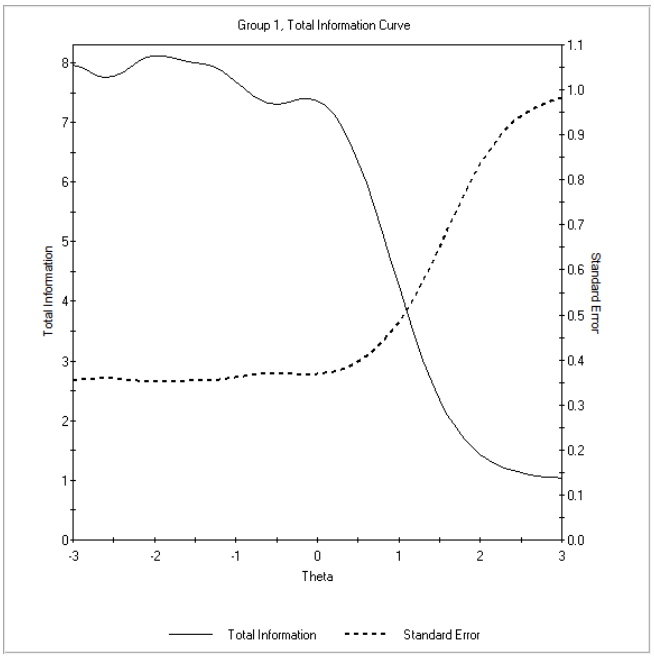
S-*X*^{2} *Item Level Diagnostic Statistics for Scale IBM*

Item	X2	d.f.	Probability
BEYMIN1	18.34	15	0.2447
BEYMIN2	42.72	20	0.0022
BEYMIN3	31.58	17	0.0169
BEYMIN5	24.02	16	0.0889

In the item level diagnostic test of Scale IBM, the p-value of BEYMIN2 was smaller than 0.01 which means the parameters of BEYMIN 2 in this scale did not fit the model well.

Figure 6

Total Information Curve for Scale IBM

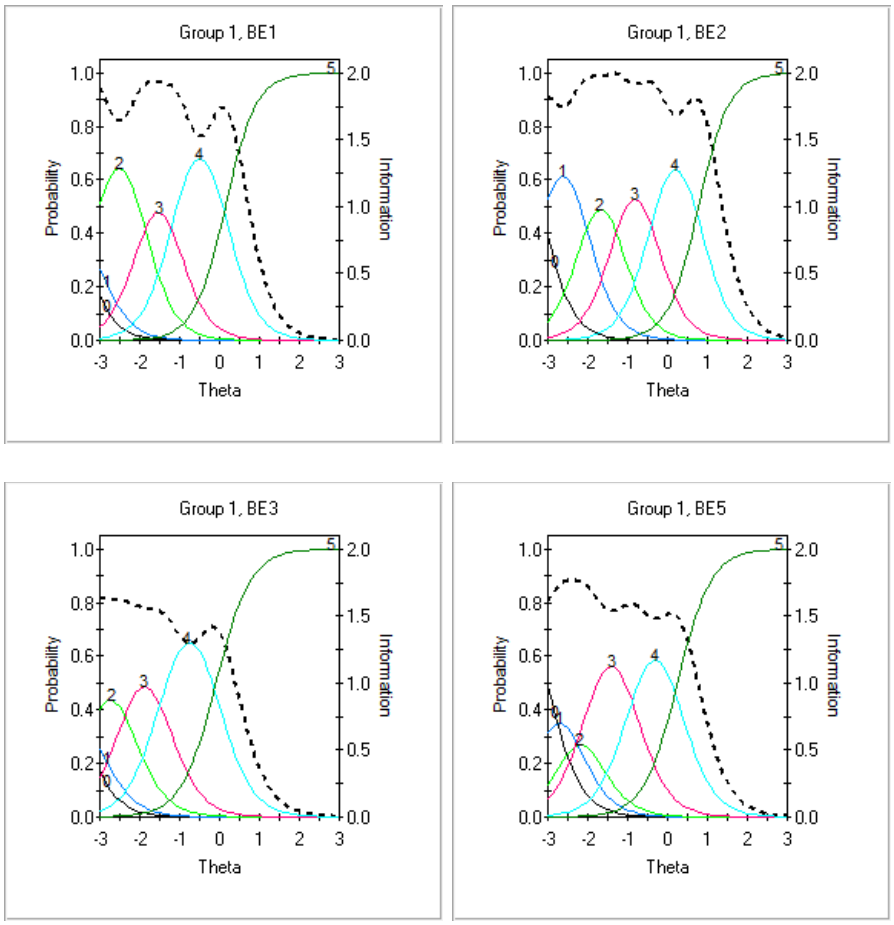


The total information curve of Scale IBM was comparatively flat trend. It dropped a little bit from -3 to -2.5, then rose from -2.5 to -2, dropped a little bit from -2 to -0.5. After θ=0, it continuously dropped. The information functions reached their maximum at approximately θ=-2. When θ=3, the information function curve was above 0.1, bigger than the Scale IWM.

From the combined trace line and informative curves, all four items provided information greater than 0.8, which means the four items were very good. From the trace line, the four items were normal, the 0 line and 1 line which means “strongly disagree” “disagree” of the item BEYMIN 1 and BEYMIN 3 were low and close to each other, this indicated the participants showed seldom negative opinions in these two items.

Figure 7

Combined Information Curve and Trait Links for Scale IBM



The Scale GC was special, it contained only two items, there was no With Item Deleted Coefficient α for this scale.

Table 40

The Summarized Statistics of Scale GC

	Response		With Item Deleted	
Item	Average	Std.Dev.	Item-Total Correlation	Coefficient α
LE1	2.961	1.878	0.9299	
LE2	3.057	1.903	0.9299	

The average values of these two items were lower than the previous items. The averages of the response were 2.961 and 3.057, the standard deviation values were larger than the items in the Scale IWM and Scale IBM. It meant the responses of these two items were decentralized.

Table 41

Graded Model Item Parameter Estimates for Scale GC, logit: a(θ - b) *(Back to TOC)*

Item	a	s.e.	b1	s.e.	b2	s.e.	b3	s.e.	b4	s.e.	b5	s.e.
LE1	6.64	1.03	-1.07	0.07	-0.58	0.06	-0.33	0.06	-0.01	0.05	0.45	0.07
LE2	46.11	0.97	-1.00	0.12	-0.55	0.52	-0.37	0.63	-0.21	0.05	0.40	0.14

The table of parameter estimates presented the parameter estimation of Scale GC. The item discrimination parameter *a* of LE2 was much larger than LE1 and quite abnormal, the standard error of b3 is 0.63 which is much higher than other parameters. The LE2 didn't fit the model well to estimate the parameters.

The item level diagnostic level was good from the table values.

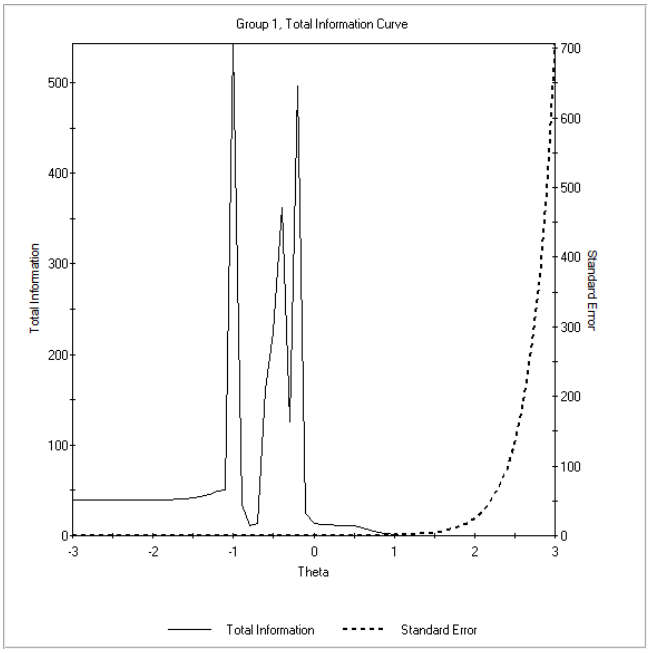
Table 42

S-X^2 Item Level Diagnostic Statistics for Scale GC

Item	Label	X2	d.f.	Probability
1	LE1	6.01	7	0.5401
2	LE2	4.90	7	0.6730

Figure 8

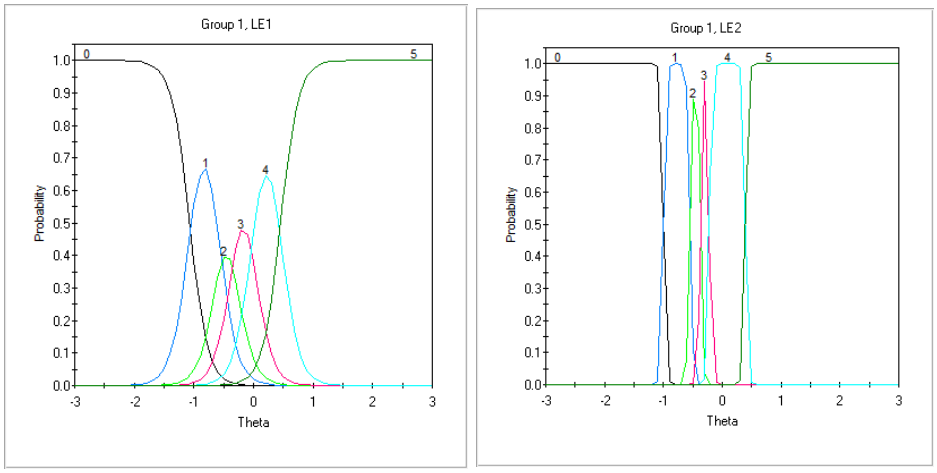
Total Information Curve for Scale GC



The total information curve was quite different from the previous scales, it was in a very weird shape, it dramatically went up and downs with three peak points. From the parameters estimates of the two items, this abnormal curve may be resulted from the item of LE2.

Figure 9

Combined Information Curve and Trait Links for Scale GC



The combined trace line and information curve proved the guess in the total information curve and parameter estimates. The item LE2 was not normal curve, and the item LE1 curve was different from the items in other two scales. The 0 and 5 lines of the items were very high, the opinions of the participants were polarized to “strongly disagree” and “strong disagree”, which resulted in big standard deviation.

This scale did not have a normal information curve. Therefore we concluded that it is not a good scale in this instrument. The LE2 was not a very good item in the model, and requires more attention from a statistical, rather than theoretical, perspective. However, the decision to use these items was to collect additional data before making a decision to revise or drop them because these two items remain very important in theory and practice.

Table 43 presents the summarized statistics of the Scale ES, the items in the Scale ES were four points Likert scale questions. The average values were smaller than the items in the previous items.

Table 43

The Summarized Statistics of the Scale ES

Item	Response Average	Response Std. Dev.	With Item Deleted Item-Total Correlation	With Item Deleted Coefficient α
LE3	1.604	0.927	0.3951	0.6832
LE5	1.300	1.021	0.3774	0.6921
LE6	2.163	0.900	0.5451	0.6361
LE4	2.335	0.766	0.5342	0.6469
LE8	2.163	0.880	0.3576	0.6935
LE9	1.661	0.975	0.4634	0.6617

The average of responses in the Scale ES were range from 1.300 to 2.335. The Item-Total Correlation were much lower than the Scale IWM and the Scale IBM, but they were all greater than 0.03, not very good but still acceptable.

Table 44 is the table of the parameter estimates of the Scale ES. The *a* values were range from 0.95 to 2.11, the range was a little large since they were 4 categories Likert scale items. All the items in the Scale ES. The parameter b_3 of Item LE4 is quite close to 0, category 4 in this item has small affect.

Table 44

Graded Model Item Parameter Estimates for Scale ES, logit: a(θ - b) *(Back to TOC)*

Item	a	s.e.	b1	s.e.	b2	s.e.	b3	s.e.
LE3	1.14	0.20	-2.18	0.35	-0.15	0.14	1.52	0.26
LE5	1.05	0.19	-1.17	0.22	0.34	0.15	1.99	0.34
LE6	2.11	0.40	-2.36	0.30	-0.81	0.12	0.11	0.10
LE4	2.02	0.37	-2.83	0.41	-1.38	0.18	-0.01	0.10

LE8	0.95	0.19	-4.04	0.79	-1.30	0.27	0.24	0.16
LE9	1.24	0.21	-1.88	0.28	-0.27	0.13	1.25	0.21

Table 45

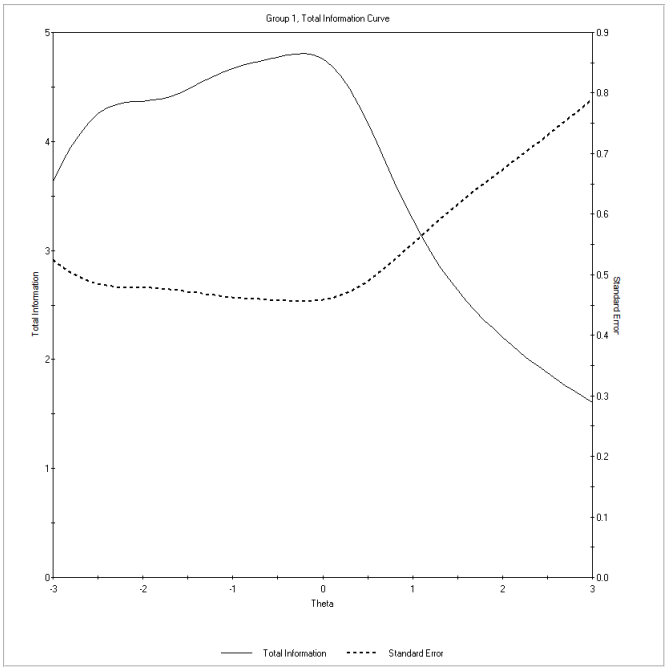
S-X2 Item Level Diagnostic Statistics for Scale ES

ITEM	X2	d.f.	Probability
LE3	31.42	30	0.3963
LE5	26.35	27	0.5005
LE6	15.10	21	0.8186
LE4	17.99	20	0.5892
LE8	21.29	26	0.7276
LE9	40.54	28	0.0589

All the probabilities of different items are greater than 0.01 which means all the items fit the model well in the item level diagnostic statistics.

Figure 10

Total Information Curve for Scale ES

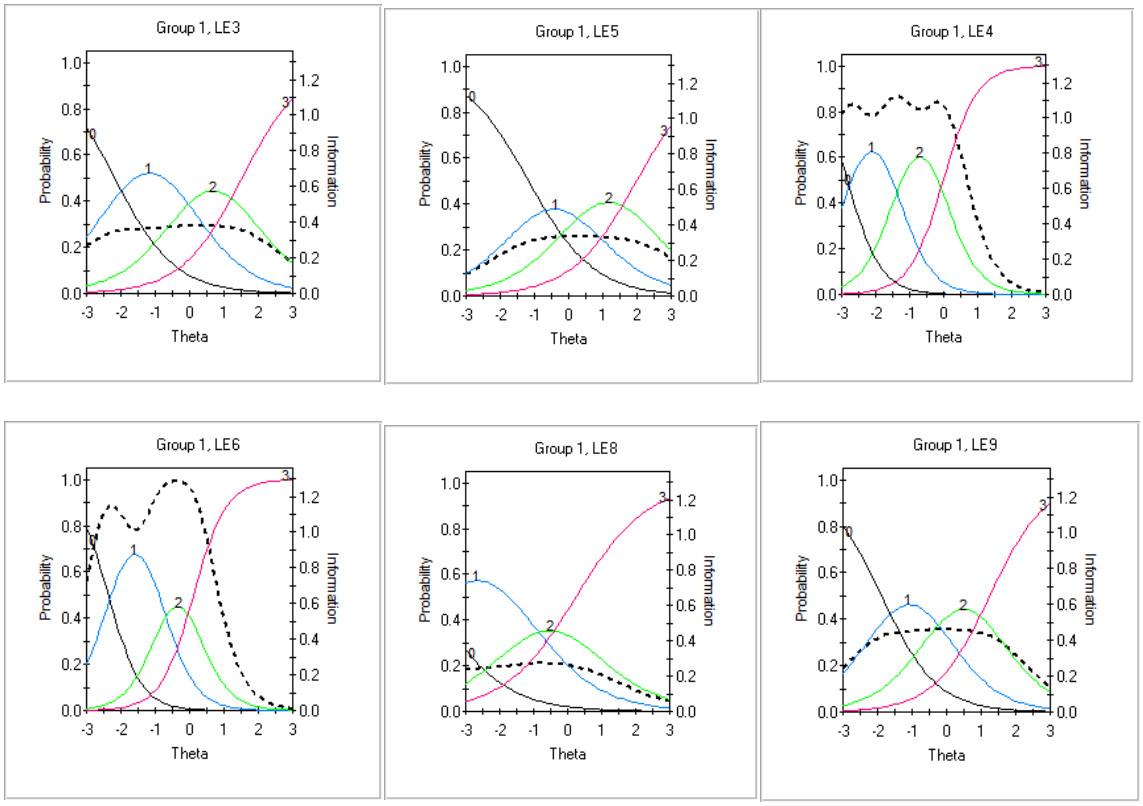


The total information curve of Scale ES started to rise from θ=-3 until θ = -0.5, it reached its peak. After the peak point, the curve s continuously dropped off rapidly after the. When θ=3, the information function curve was above 0.3.

From the charts, we could clearly see that this scale fitted the model not as good as the Scale IWM but better than the Scale GC. But for the combined lines, LE3, LE5, LE 8 and LE9 presented the information less than 0.5, the trace lines of LE3, LE5 and LE8 were flat, especially the item of LE8. The items had no big problems, but they did not distinguish the answers well. The informative lines of LE4 and LE6 were normal.

Figure 11

Combined Information Curve and Trait Links for Scale ES



Answers to the Research Questions

The quantitative data analysis provided the instrument level analysis and item level analysis for the measure development. It consisted primarily of summarized data of descriptive statistics, correlations between item and the instrument, chi-square test, and a variety calculation in IRT models, and IRT plots. The descriptive statistics calculated item response means and standard deviations and standard errors of parameter. From the data analysis, the researcher can answer each of the research questions proposed in this study.

RQ1: Can we develop an instrument to measure a student learning outcome which is called integrative learning in a thematic general education program?

The data analysis shows that it was possible and practical to develop a measure to evaluate student learning experiences and outcome from the innovative general education program. The process validated the framework that resulted from the initial literature review but found the three dimensions proposed in Chapter 3 did not support the assumption of unidimentionality in the IRT models. The researcher then used exploratory factor analysis to further examine the third dimension called “Integrative Learning Environment.” The factor loading output suggested this dimension could be divided into two scales, subsequently named “General Climate” and “Environmental Support” by the researcher. The factor structure was changed to four scales.

The reliability within Classical Test Theory of the whole instrument was over 0.9, which means the construct of instrument was very reliable. Then the researcher calculates the reliability of each scale. They were 0.9484, 0.8430, 0.9639 and 0.7162, all of them

were acceptable. The highest reliability of the internal construct was the Scale IWM and Scale GC. Then the research started to analyze the data at item level. Through the summarized statistics, parameter estimates, IRT plot combined trace line and information curve, most of the developed items were fitting the IRT grade response model well. The parameters of several items were not as good as others from the statistical perspectives.

For example, the parameters and IRT plot of LE2 were not normal, the item was significantly lacking the fit to the model. But this did not mean they should be absolutely deleted from the instrument, because the quantitative data of LE2 was not very good. was asking about the general environment of Pathways minor, it was very important for the practice, it may need a revision of wording and accumulate more data to revise or delete in the future. This will be discussed in Chapter 5.

No instrument can be perfect. Although some parameters of few items were not perfect, overall, the instrument was reliable and consistent from the statistical perspective.

RQ 2: Is there a measurement model that best fits the instrument?

Yes. The current study compared the model fit indices of four scales: Integration within Minor, Integration beyond Minor, General Climate and Environmental Support, of a Grade Reponses Model against three other IRT models (Partial Credit Model, Graded Response Model, General Partial Credit Model and Nominal Response Model) with an Exploratory Factor Analysis designed specifically for the supportive environment section to divide the original factor into two. The researcher compared the fit indices of AIC, BIC and -2ln Likelihood. Finally, all the indices of the Grade Reponses Model were

smaller than other models', this indicated the GRM fits the scales better than other modes. The researcher decided to estimate the parameters of each item by GRM.

Then the research presented the table of parameter estimation, the table of item level diagnostic statistics, plot of total information and trace line to analyze each item. The parameter estimates showed normal numbers of most items, few items needed to be revise. The main problem for the parameter estimates were small effect of some categories.

From the quantitative analysis, the research provided sufficient evidence that the instrument is practical for the program, item LE2 needs revision, and item THEME 2 may require additional attention in future analysis.

Conclusion

This chapter was the main body of the item analysis in this research. In this chapter, I presented the results of qualitive and quantitative analysis process and outcomes in details.

The data were analyzed at different levels in multiple research methods. The results of the analysis responded to the research questions posed in the study. The qualitative method used to validate, revise the framework established from the literature review and lead to deeper insight into integrative learning bring in additional items on the instrument. The first research question examined whether it is possible to develop an instrument to measure a student learning outcome which is called integrative learning in a thematic general education program. Data support the efficacy of the instrument showing it reliably elicits data that can used to evaluate innovative general education programs.

This is further supported through a graded responses model that was applied to the instrument and demonstrates strong item analysis useful to decisions at the item level.

In the following chapter, the researcher will briefly reintroduce the purpose of the study, summarize findings, and present implications for practice and research. Limitation of the research will also be presented.

Chapter 5 Discussion

This research developed a measurement instrument to assess student perception of integrative learning in a novel academic minor program at a public research university in Southeastern United States. The "pathways minor" in unique in that general education is shifted to thematic courses intended to help students integrate learning across disciplines. The program encourages students to simultaneously acquire general knowledge typical of many general education curriculums while also exploring deeper, integrated learning experiences applicable to a real-world area of focus or theme. The theme of any such minor program consists of required foundational courses, elective courses, and a project-based capstone course. Instead of learning fact-based knowledge from disconnected areas, students in the minor have an opportunity to integrate academic areas through an integrating theme. Diverse instructors from a variety of academic areas teach and facilitate learning experiences within each minor.

Since the format of the so called "pathways minor" is innovative, the degree to which students demonstrate an ability to integrate knowledge becomes the priority of various stakeholders including faculty, university administrators, and students themselves. This study developed an instrument designed to measure this ability and provides a tool for new insights to integrative learning. While the main outcome of the research is a new assessment tool for a unique, innovative general education program it has implications for similar programs in higher education. With its special characteristic of both interdisciplinary learning programs and thematic general education, this research

creates a solid foundation to assess student learning outcomes and evaluate program objectives.

This research started with a traditional literature review to establish an initial framework for the measurement. The initial instrument was evaluated by instructors, experts, and stakeholders, then it was distributed to students who were enrolled in the capstone courses of the pathway minor program. Data was collected in in fall 2020, spring 2021, and spring 2022. The data collection was completed after one wave of a pilot study and two waves of a field study. Then the research adopted Classical Test Theory to calculate the reliability. Next, IRT models were applied to examine the instrument from both the instrument and item levels. The official proposals for minors submitted by faculty and the open-ended questions in the survey completed by students were considered as complementary qualitative data to establish the framework and revise the instrument.

The process of developing the instrument gave the program a practical way to conduct program evaluation, and revealed the cognitive development of how students integrated knowledge as well as the crucial factors that affect students integrative learning. While not central to this study, findings showed, for example, that connections across areas and reflection in learning were key learning outcomes among students. The results of this study suggest the instrument can be reliably applied in future learning assessment and program evaluations. Additionally, the assessment results proposed practical suggestions for the pathways minor program and similar programs at other institutions of higher education.

Summary of Findings

This study reviewed historical and theoretical literature that revealed the importance of innovation in general education programs and situates integrative learning in general education in a modern era of higher education. The researcher organized related literature into several topics including the history of general education, integrative learning process, and the development of student learning assessment. This was a critical step in this research because it not only contextualizes emerging general education programs, but it also provides important insight to developing the instrument that resulted from this study. Users of this instrument will benefit from this as a way to support a transition from traditional general education programs to more progressive approaches to foundational curriculum. While the historical roots of general education remains deep in American higher education by national and local traditions or convention, innovation is inevitable.

The history of general education reminds us that such programs have their own targets of exploring the world around them, developing intellectual expertise, and personal growth. One central goal of higher education is to produce graduates who are capable of dealing with uncertain and unprecedented times, but moreover to be productive citizens. A good general education needs students both widely and deeply involved, reflecting their learning experiences, and meet a range of learning aspirations. These special characteristics of general education provided basic direction when developing any instrument designed to assess the educational objectives of general education. For example, this study took the idea of applying knowledge to real context as

critical to the concept of integration. Similarly, the item that connects students' learning to professional development casts students' perceptions of this learning experience into the future. In short, there is a sense of relevance and usefulness of the knowledge.

Importantly, the literature review revealed the legitimacy of applying an interdisciplinary approach to promote integrative learning (Hartesveldt & Giordan, 2008; Klein, 2005). Further, we know that conditions must exist that foster an interdisciplinary approach that results in integration as a critical learning outcome. The literature review further revealed the context of integrative learning and general education which were considered as two different learning formats in the past, suggested integrative learning was perfectly implemented in a theme based general education program now and demonstrate thematic interdisciplinary programs may be an ideal environment to promote students' integration in their learning and provide theoretical foundation for the pathways minor practices.

Most importantly, the review revealed a gap in assessment tools and practices that can be reliably used to measure integration. However, prior studies have uncovered important aspect of integration that were used in the construction of items for the proposed new instrument. Several studies distinguished different stages of integrative learning happens within an interdisciplinary course setting (Jacobs, 1989; Repko, 2008) and synthesizing knowledge from different courses could start from an interdisciplinary program during the learning process (Jacobs, 1989; Newell, 2001). Importantly, integration is hierarchical. Higher levels of integration always involves a pre-requirement (Bell et al., 1995; Linn, 2000). The existing literature demonstrated a solid disciplinary

foundation of knowledge as the first step for students to integrate. All levels of integration started with a well-established understanding of knowledge within a discipline since disciplinary grounding avoid the risk of the linked courses being "a mile wide and an inch deep" or "a light educational experience" (Mansilla & Duraising, 2007). Suffice it to say, the initial step of integration happens *within* a discipline. Only then can students start to understand knowledge in an integrated manner because of the resulting disciplinary connection points. Without sticky connection points, students' frame of reference is muddled and likely integration across disciplines is difficult if not impossible.

Where a general education curriculum is not fixed on integration from the beginning, it may be more likely for students to integrate by chance rather than by design. Course construction, therefore, has to meet certain requirement of integration to help students integrate. In such cases, the course can emphasize the deepness and continuousness of knowledge to set a consistent knowledge base for students. Interestingly, the next-level integration happens among disciplines based on the foundation of disciplinary knowledge. Then in one final step, and perhaps most critical, the application of knowledge must be made to a real-world context including real problems and previous experiences.

The detailed process of integration found in the literature inspired and helped the researcher to establish the new framework for the assessment instrument developed for this research. The framework was strictly built on the study of the previous research and literature, and all the items were firmly evaluated by experts. The study does not treat

integrative learning as a static result and it uncovers the cognitive process of building integrative learning. The study defines integrative learning as a complexity with hierarchical levels and tried to establish relationships between learning levels. The learning outcomes can be affected by both inside factors of the program such as the curriculum, the first level of integration in knowledge integration. This level requires the construct of the curriculum in any program to be related to a central theme and to promote integrative learning for students within the courses. To integrate all knowledge across disciplines they have learned in the program, students have to master the requisite knowledge grounded in each discipline, then integrate them in the academic area and try to solve problems from diverse perspectives. After this process, integration comes to a higher level that signifies a leaner-initiated integration. It means to connect, apply, and synthesize learning across the learning and real contexts (Barber, 2020) which needs students to integrate knowledge with their own lived experiences and in real-world scenarios. In this level of integrative learning, learner-initiated connections emerge between their life experiences, including their education, through a sort of reflexivity on their experiences including learning, career goals, and knowledge acquired from their major. Since, integrative learning also closely related to the learning environment, the framework including perceptions of the learning environment.

Next, the research comes to an empirical undertaking that resulted in the development of a new assessment tool designed to measure student perceptions of integration. Through examining both qualitative and quantitative responses by student participants, the instrument provides a valid and reliable measure to assess students'

integration in an interdisciplinary general program. Through a process of developing an instrument based on its own advantages and deep investigation of experiences, accrued evidence led to important revisions and subsequently supported the usability of the instrument.

In the qualitative analysis, proposals of different minors were reviewed to understand proposed core learning goal of each. Together with the literature review the proposals provided insight to build the initial framework. Then the open-ended questions collected from students in the initial two of the three waves of data were analyzed. This research adopted only the first two waves, because the qualitative data was mainly used in the process of building framework. After the analysis of data in first wave in spring 2021, the framework was fixed. This final framework consisted of three dimensions: Integration within Minor, Integration beyond Minor and Integrative Environment. Combined, these dimensions are an effective means to assess the integrative learning in the program.

Next the field study used the proposed framework as a way to identify theoretical key words to analyze the data. The result of this analysis was three element of the instrument that needed to be improved: reflection, innovation, and course construction. Based on this, the researcher added one question on vertical construct of the courses which means the continuousness of the course difficulties to the Internal Construct in the integration within minor dimension, a question on self-reflection and a question on innovation to improve the instrument. The qualitative data supported the effectiveness of the instrument framework, the students' answers to the open-ended questions helped to

established and revised the framework, then showed that the scales in the instrument were pragmatic.

The Classical Test Theory and Item Response Theory were both used in the data analysis process. The two approaches have their advantages to analyze the instrument data. The different methods to calculating the reliability of the instrument were all over 0.9, the instrument had a very good reliability. Then the researcher used confirmatory factor analysis loading to assess the one-dimensionality of each dimension set up through literature review. The percentage of Eigens presented Integration within Minor and Integration beyond Minor were unidimensional, but the Integrative Learning Environment factor was not significantly unidimensional. Next, the researcher used exploratory factor analysis to explore the Integrative Learning Environment dimension, the loading suggested to divide the factor into the two scales, the researcher called them "General Environment" and "Environmental Support", two items LE1 and LE2 belonged to the General Environment, the other six items were clustered to the second factor. The Classical Test Theory was used after the confirmation of the structure of the instrument to test the reliability of the whole and each scale, the output showed a good consistency of the inner construct of the instrument. This means the framework provided an effective perspective to assess the students' integrative learning in this interdisciplinary general program. Then compare model fit to use of the polytomous IRT models for trait score estimates. The item analysis compared the traditional indicators indices of likelihood ratio, AIC and BIC indices of the item through Partial Credit Model (PCM), Graded Response Model (GRM), General Partial Credit Model (GPCM) and Nominal model

(NM), the output suggested the GRM was the best IRT model to fit the data in this research. The item analysis was based on GRM which consists of traditional summarized statistics, parameters estimation and IRT plot including the informative curve and category response curves. The item information exhibited the evidence of making decision to keep or revise the items to make the instrument get more meaning from the statistical perspective.

The calculated parameters of the most items were good, but some items were not as good as others. In Scale of IWM, the discrimination was very good. When deleting no matter which item, the coefficient decreased, which means the items were all important to the scale, the THEME 2 differentiated among trait levels worse than other items, and it provided less information than others, but the data couldn't prove the THEME 2 was not good enough. The indices suggested the Scale of IBM also has a very good discrimination, the items in this scale showed normal indicators. The outputs of the Scale GC were not as normal as others, all the plots were up and down dramatically because the LE2 didn't fit the model well. The plot shapes of Scale ES were normal, the Item-Total Correlation were much lower than the scale one and the scale two, but still acceptable. All the results showed the framework was effective to assess the learning outcomes in the Pathways minor from the statistical perspective, but several items for example THEME2 and LE2 needed to pay attention to. The different plots produced by IRT models visualized the parameters and gave a direct estimate of each item. From the informative curve, we can see some items provide more information than others while others can offer more information. The shapes of the trait lines showed students' attitudes towards

the questions, which category they prefer to choose. When the line is low and narrow, it means the student didn't like the answer. It can also affect the quality of the item, whether they provide leaning responses. The plots in this research were quite align with the parameter estimations.

Both of methods provided substantial evidence in the process of developing the instrument. The analysis offered support that the framework established from the literature was practical. However, the results also proposed some necessary improvement. For example, adding additional questions to the course construct and reflection, and adding innovation improved the instrument.

Overall, the instrument that results from this through examination of the literature combined with substantial empirical evidence is reliable, valid, and importantly, practical. The instrument will can be used to collect data as a part of program evaluation and can also serve as a foundation for other interdisciplinary programs to develop instruments that are locally responsive.

Discussion

This research proposes a new measurement tool to assess student perception of integrative learning in an interdisciplinary general education program. The research set up a solid foundation and it brings important implications for researchers and assessment practitioners, which will be discussed later in this section. The results of the study indicate that the instrument is a valid and reliable measure of integrative knowledge. The central outcome of this study is a new assessment tool that can be used by researchers and

assessment practitioners in higher education. Further, it can be used by stakeholders at the university were this study was conducted to improve the pathways minor program.

For the instrument and its related practices, we can develop further research and practices to improve student learning. In the following section, I will propose how researchers can use this instrument and suggest new studies to refine the instrument and test it in different contexts with different student types and at different types of colleges/universities, and then examine the methodological critique of the research. Next I will explain how practitioners can use this instrument to inform integration in general education programs. Finally, I will state how the results of this instrument should be ethically used as one data point in decision-making. All these discussions can reduce the limitations of this research in the future.

Implications

Implications for Future Researchers

I propose several directions for future research. These include two directions: (1) research on the instrument itself and (2) using the instrument developed to study a wider range of students and academic programs.

The instrument was tested using two waves of students. Arguably, it would be valuable to collect and accumulate data and develop it into a longitudinal research design. Given that a larger sample size and more precisely sample power is of principal concern when designing a study (Tabachnick & Fidell, 2013), sufficient statistical power will improve the possibilities of observing true relationships in the dataset (Wolf et al., 2013).

With the expansion of the sample size, the quantitative model will continue to increase its power, may bring more information for the researchers to improve the instrument.

Then the researchers might be able to build more complicated statistical models such as structural equation models or hierarchy linear model to investigate factors that affect integrative learning. The factors concluded in this research, mainly from the literature review, that the final instrument is useful for understanding the efficacy of integration in a general education program. In future studies, a more nuanced approach is necessary to understand differences among special populations of students such as minoritized and underrepresented groups. Or the effects of different pedagogy or features of a class like number of students. These questions can be answered in different statistical models, for example, hierarchy linear model can explore components to influence integrative learning in multiple levels including personal and organizational dimensions. Organizational types or size, and organizational climate from a collective level can be investigated in the future's study. The process will absolutely bring theoretical meaning to the field. Then researcher can gather the impact from various factors, then revise the instrument, finally broaden the use of instrument into specific contexts.

Because of the timeline limitation, this research adopted the text analysis of students' comments which were very good complementary materials to quantitative models, more qualitative data can be applied to this research to refine the instrument. Learning is not a snapshot but a long process, as I mentioned before, qualitative method is an appropriate way to build new theory and deeply investigate a complex process (Graebner et al., 2012), to further interpret students' individual experiences in the process

of integrative learning, qualitative method has its natural advantage. In the future, many ways can be explored to extend the qualitative method. In addition to open ended questions, group interviews, individual interviews, field observations are useful approaches to collect qualitative data of students learning. As a researcher, merging into the data, understand their behaviors and learning habits will lead to a constructive vision towards students' experiences. Different qualitative analysis ways also create unique intuition upon the same data, grounded theory evolves data in an inductive generation of theory derived from data (Walker & Myrick, 2006) while phenomenology is often regard as pivotal to the descriptive or interpretive paradigm (Clark, 1998), researchers may emphasize more on theory building to improve the framework, instructors in the front line of teaching may pay more attention to meaning of phenomena in students learning.

From a purely measurement perspective, there are several paths for researchers to consider in the future. First, at this stage of development the instrument is fixed on norm-referenced scores. Given that we only have set a general standard for the whole instrument, we only know that higher score means higher integrative learning. The current study did not set a criterion to determine "good" integrative learning in each scale or the whole instrument. Future studies might explore a target score of the instrument with a cut-score, or criterion referenced score, for the instrument. Doing so would allow users of the instrument to make more substantial claims based on student scores.

This research provided a unique score for each subscale. Reliability evidence was provided for each and provides strong support for the instruments use. Whether we can provide a *total score* of integrative learning remains to be determined and is a source of

future research. The combination of a total score and subscale scores would be useful to researchers and educators in the field.

Importantly, this research provided both CTT reliability and IRT item analysis. The results of this analysis have important implications for researchers and educators. From a practical standpoint, a 6-point Likert scale is easier to understand and more accessible for stakeholders who don't have knowledge in measurement theory. But for researchers, the IRT models provide a more nuanced understanding of the instrument that can be used to revise the instrument. IRT models have specific processes for verifying item fit that allow the identification of items that do not meet the assumptions of the model (Embretson & Reise, 2000). This information is more precise for determining whether test items are operating according to the primary model.

To apply the instrument to a broader context and develop more theoretical meanings, researchers can extend the boundaries of respondents and apply this instrument to new studies and test it in different contexts. This is an instrument for college students in the program. It would be helpful to collect and analyze data from instructors as well, then compare the model fit between the data collected from different sample groups, this process will iteratively validate data from different respondents. The data from instructors can also show their understanding of integrative learning, broaden integrative learning from their perspectives. When comparing the perceptions of students and instructors, the results would be used for better practice in the future. For example, assess different student types and at different types of colleges or universities other than this research. It will benefit to break the limitation of the instrument which was developed from a certain

group of students. With more participants from different backgrounds, we will know more limitations of the instrument, at the same time the instrument can be updated to contribute more to the theory of integrative learning in different contexts.

Although this is an instrument for an interdisciplinary general education program at a specific university, this instrument emphasizes very general learning outcomes of integrative learning which can be applied to other similar contexts. In my research, the pathways minor emphasizes integrative learning of general education under a common topic. For other learning formats, for example learning groups or first year learning program, researchers might highlight a specific stage of integration relevant to the goals of the program.

In this research, I regard the integration in different levels as the same importance. The researcher can dive into the integrative learning process to investigate the weight of integration stages in the integrative process. In this research, I listed several levels of integration, from knowledge grounding to application. Which stage is most important for students in their experiences as an undergraduate student? Can we explore more in the most important stage of integration? This process probably led to more theoretical meaning to the field deduce the limitations of the using the instrument in a special context.

The entirety of this research based on the assumptions that all integration happens in a context of face-to-face teaching and learning. The existing literature discussing integration were mostly focused on the integration happened on face-to-face learning experiences (Bell et al., 1995; Lattuca et al., 2004; Newell, 2001). This study occurred

during a global pandemic when many, but not all, courses were offered in an online platform. Subsequently, students' interaction with peers and instructors were dramatically changed and have implications for integration. Do students integrate knowledge still in the traditional way in a post-pandemic era? The research provided a good foundation for future researchers to understand integration as it happens in different learning environments, but it is open to more examinations for improvement. It is likely that additional items are necessary to both capture virtual learning spaces and perhaps new components of integration that are unique to online courses of study.

Implications for Practitioners

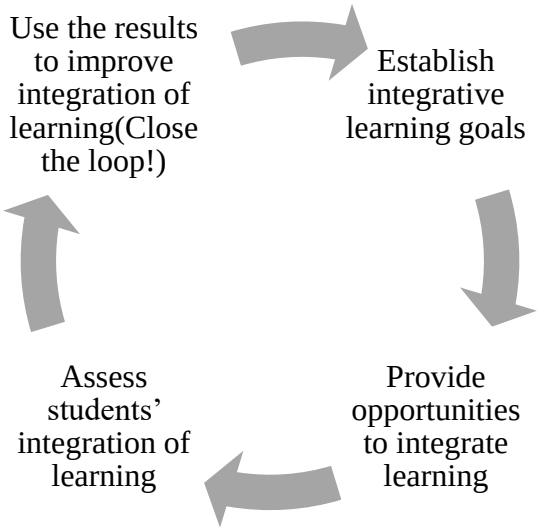
The theoretical research and the practice of assessment are reciprocal processes. By this I mean that researchers provide insight for practice and then generalize theoretical inspirations from the real practice of assessment. There are essential discrepancies to think through between research and real assessment practice (Barber, 2020). For example, the assessment cycle shown in Figure 12 was proposed by Barber and creates a link between what researchers and practitioners independently consider in their work.

The work of this dissertation results in a measurable framework that presents the integrative learning process of students and established goals of integrative learning. Including the integration of knowledge within and across the disciplines. The highest-level learning outcomes of integrative learning in an interdisciplinary general education program is integration with the real-world context and students' own experiences. This study considered the characteristics of a unique program and prior research to design an instrument that measures integration. To date, no such instrument exists in an assessment

practitioners' arsenal of tools to document student learning outcomes. Results of this study suggest the factors that should be assessed and the essential elements for such an assessment instrument.

Figure 12

The Integration of Learning Assessment Circle



Note. Facilitating the Integration: Five Research-Based Practices to Help College Students Connect Learning Across Disciplines and Lived Experience by Barber, James P. Stylus Publishing. COPYRIGHT © 2020 BY STYLUS PUBLISHING, LLC.

The new assessment instrument now bridges the gap between theory and program practices where integration of knowledge is paramount. Through empirical evidence of integration, in this case student *perceptions* of integration, we are closer to unlocking the mechanisms that are central to high impact learning experiences. Of course, this instrument can be used to improve such program including course construct, instructional quality, and producing evidence for quality assurances such as accreditation. It becomes

essential for practioners to make data useful to and used by faculty members. Perhaps the most essential step in assessment is to "close the loop" and make the assessment itself more meaningful. Meaningful assessment tools such as this one can be used by faculty to make decisions about individual courses, curriculum, and academic programs. Similarly, academic administrators might use it to look across programs to investigate best practices.

The final goal for this research is to build a platform for practitioners to improve student learning, namely in the area of integration which is new and to date has been elusive. The creation of a new instrument will result in higher levels of engagement of stakeholders to actively participate in improvement of student learning as evidenced in data originating from this assessment tool. Directors, instructors, and students all have different perspectives regarding student outcomes such as integrative learning. By directly measuring integration, student perception of their experiences can be used to inform ongoing development and improvement of similar high impact practices in higher education.

However, as experts use this instrument to assess learning an improve programs, interpretation of the instrument will become a central concern. What does each number mean? Is a 5.5 to 6 "good" or "good enough" for a program? How can or should practitioners interpret the assessment results? Here, common sense and best practices in assessment should remain at the forefront of interpretation. For example, assessment results should involve a range of stakeholders including instructors, program leaders, and students themselves. Further, use of the instrument and interpreting results must be a

local decision. The mission of the institution and its educational priorities will inform curricular development. Integrative learning, for example, may differ across institutions and programs. Contextualizing data that results from this instrument must result from discussions among key stakeholders. Only then can the data become meaningful in a practical sense. This may be especially important in emerging programs such as the one used in this study where it is largely experimental and does not have a long history to base interpretation of data that leads to quality improvement. Stakeholders will need to discuss, among other things, threshold points and rubrics and participate in reading the qualitative data, which can provide deep insight of students' satisfaction and understanding of their dissatisfaction. Then the assessment can set the standards and rubrics to interpret the data.

Analysis of data in this study should provide stakeholders with strong evidence that any report generated will be reliable and valid. However, what remains is contextual, detailed description and discussion of the scores. In fact, the final instrument has been adopted by the university where the study was conducted. Reports have been developed that show data visualizations and a detailed report including response distributions. The reports, while basic, contain the sample size and mean scores. These reports were sent to the program leaders after the survey closed. Additionally, text reports were provided that include more detailed information. Examples of these reports are provided in Appendix F. Again, while these reports are simple, they provide important insight to integrative learning which to date has been absent from discussions.

Appendix G shows the raw question report. It clearly showed all the percentage of each answer. The table attached to the figure explained the count, percentage, standard errors and standard deviation of the answers. Combined, these reports were sent to leaders for them to have access to student responses with the intention to use as a datapoint for discussions about courses and student experiences in those courses. The resulting reflect in changes are intended to improve pathways minor programs.

The office also used this instrument to engage stakeholders during a summer institute. The director of the program hosted a workshop for faculty program leaders where he shared the overall enrollment data regarding the minors and the overall results from the instrument. The workshop motivated important discussions on the student experience, shortcomings of the integrative minor programs, and other important aspects of the minors. In addition to this, faculty leaders were provided program-level reports that were shared with them beforehand. They then got in small groups to talk about their results and developed ideas for improving their programs based on the analysis provided. Colleagues from the university's teaching center were on hand to assist with the discussion, contributed ideas to improve teaching and learning, and answered questions. This is a dramatically good starting point for education practitioners to take because discussion was informed by data which led to important insights from student stakeholders and resulting in what is generally referred to as "closing the loop" in assessment.

Other than benefit for researcher to construct theoretical framework, qualitative method acts not only for instrument revision and theory building, the process itself can

also help students to document and reflect their own learning experiences in the assessment practices, this is an excellent way to capture and promote students integrative learning by qualitative itself. Take myself as an example, I have participated in an interview as an interviewee for research on international students' experiences that conducted by another researcher at a big public university. In the dialogue between the researcher and me, I started to reflect on the change process of my personal value since I started my study in the US, then I sort the timeline of my life, and discovered that I had many personal growths in the past. It proved the qualitative method sometimes can interactively intervene the participants experiences and thinking, make them reflect about their personal experience and adjust the direction of their future. No matter what kind of instrument or assessment, our final goal is to help students gain more from the program, the qualitative method may help students to improve their learning strategies. Based on this, we can add more qualitative interviews as methods to collect data, for example, one on one interview to invoke students' deep reflection of their learning process, or group interview to generate more constructive interactions among peers when discussing of the learning experiences. The questions will be discussed with students can be concluded from the integrative learning framework proposed in this research. In this kind of assessment practice, we can help them to brush the stages of integrative learning, get more deep thoughts on improving their learning. This is similar to a field study constructed by practitioner and participants where the interactive construction in the practice can help both practitioners and students to have reflective thinking on the program.

All the directions for future study proposed by this research can be developed to improve the students learning. Additionally, there are many other assessments methods at different levels can be applied to specific programs. If, for example, different assessment projects can invoke students to actively engage in how to integrate across disciplines, instructors of different courses have more communication on their course constructions, then this research has successfully set up a milestone.

While there are many upsides to the use of this instrument, tremendous caution is advised. I advocated several future directions for researcher and practitioners, yet we need to pay more attention to the use context of the instrument and potential for misuse. The use of this instrument should remain exclusively to assess the learning outcomes of integration, not involve informing teaching performance or aid in financial decision making. Since this is survey with general questions to gather answers from students, it is a window to their experiences and explicitly not as a performance indicator for instructors. Further, the Pathways Minor is special because it is characterized as thematic general education program. When we utilized the instrument, the first thing is to make sure the program goal is aligned with the goal of Pathways Minor. Second is to make sure the assessment level as well as assessment goal are the same with this research. The goal of project or research is always the priority to consider.

Limitations

No study will provide a limitation free research. As with any research, there are several limitations that I identified. The first dealt with the sample. Although I tried my best to promote the assessment project and used best practices to improve response rates,

the number of available students in the minor capstone course remains low. The Pathways Minor program, while growing rapidly, is new and relatively small compared to traditional minor programs.

Importantly, this study was largely conducted during the global Covid-19 pandemic. Most students in this study would have begun the program before the pandemic, but the capstone experience would have been impacted. Additionally, the pandemic likely impacted response rates and possibly non-response bias. Because response rates were initially low, I decided to collect two waves of data to ensure sufficient sample size. I subsequently combined the two waves of data collected at different time points. This improved the sample size for purposes of analysis, however, the sample size remained relatively small and should be treated with some amount of caution in any interpretation of the findings. However, while this substantially improved the number of responses, student perception of their experiences may have been influenced by both educational and personal circumstances. This was beyond control of the researcher.

Most capstone courses used in the study were changed from face-to-face format to online format. These likely changed student experiences and their perceptions of the program, especially the interactions with faculty and peers. While the study was to collect data in the spring 2020 to spring 2021, students were shocked by the pandemic and needed to adjust the psychological conditions to the situation. As we see in the descriptive data, the majority of participants were junior and senior students. This group of students were previously exposed to face-to-face courses in their first year in the

minor. However, this special situation still has an unknown effect on the assessment results.

Another limitation as that all students come from the same university who experienced a general education program unique to that institution. It is therefore possible that students at this institution differed in some important ways from students at other colleges and universities. An art university may have totally different definition of integration with polytechnical university since the climate and purposes of learning are different. As a result of the range of behaviors characterized by the concept of integrative learning, the assessment of student outcomes has to be to defined by a particular university or program truly expects students to do as a learner (Ross Miller, 2005). Although this did not affect the immediate purposes of this study to generalize the framework of integrative learning in general education, it might have influences on future studies using this instrument when university or program define what is integrative learning for itself, the assessment tools are offered as potential models for modifying, not easily adopting because it must apply its own climate. This is a limitation that can be solved if more than different types of institutions or programs participate in this assessment project. But this moment, the sample of student coming from a unique university remains to be a limitation.

Another limitation is the researcher stopped working at the Office of Undergraduate Education during the operation the assessment process. Based on the special situation, the researcher had to transfer the data management to the graduate assistant when collecting the data in spring 2022. Because of this, the researcher did not

participate in this data collection and cannot fully account for the process of the data collection during this wave. The platform in this wave was also different from the first two waves and data cleaning may have varied. For example, QuestionPro does not provide how unsubmitted survey data was treated and it is unclear how the software handled the uncompleted data of participants. Although this may not affect the final statistical results, the method to deal with data may have been different than the previous waves.

Despite these limitations, this study is still worthwhile and provides an important and much needed tool to understand student learning experiences, namely their perceptions of integration of learning. The instrument has been adopted in the real practical work, it reveals how students continuously integrate knowledge by providing a critical initial step to create empirical evidence of student learning in a progressive general education program. Further, it sets up a milestone for future research on students' learning outcomes of interdisciplinary program in different contexts.

References

Arthur R. King and John A. Brownell, The Curriculum and the Disciplines of Knowledge: A 7heory of Curriculum Practice. New York: John Wiley and Sons, 1966.

Association of American Colleges. (2004). Taking responsibility for the quality of the baccalaureate degree. Assn of Amer Colleges.

Association of American Colleges and Universities. (2004). A statement on integrative learning. Retrieved December 4, 2009, from http://www.aacu.org/integrative_learning/pdfs/ILP_Statement.pdf

Association of American Colleges and Universities. (2008). New leadership for student learning and accountability: A statement of principles, commitments to action. Association of American Colleges & Universities.

AAC&U. (2007). *College Learning for the New Global Century*.

AAC&U. (2022). *No Title*. https://www.aacu.org/initiatives/value-initiative/essential-learning-outcomes

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

American Council on education. (1949). *Wanted: 30,000 Instructors for Community Colleges: A Bulletin for Prospctive Teachers and for Institutions Preparing Teachers*. American Council on education.

Anderson, T. (2003). Modes of interaction in distance education: Recent developments and research questions. Handbook of distance education. In *academia.edu*. http://www.academia.edu/download/3239267/Leading_ODL_Futures_In_the_Etern

al_Triangle_The_Mega-University_Response_to_the_Greatest_Moral_Challenge_of_Our_Age.pdf#page=155

Andrich, D. (1978). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, *2*(4), 581–594.

Ashby, I., & Exter, M. (2019). Designing for interdisciplinarity in higher education: Considerations for instructional designers. *TechTrends*, *63*(2), 202–208.

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel*, *25*(4), 297–308.

Astin, A. W. (1993). *What Matters in College*. 398.

Asyari, M., Al Muhdhar, M. H. I., Susilo, H., & Ibrohim, I. (2016). Improving critical thinking skills through the integration of problem based learning and group investigation. *International Journal for Lesson and Learning Studies*, *5*(1), 36–44. https://doi.org/10.1108/IJLLS-10-2014-0042

Austin，, A. W. (1985). Achieving educational excellence. *Francisco: Jossey-Bass.*

Awang-Hashim, R., Kaur, A., Yusof, N., Shanmugam, S. K. a/p S., Manaf, N. A. A., Zubairi, A. M., Voon, A. Y. S., & Malek, M. A. (2021). Reflective and integrative learning and the role of instructors and institutions—evidence from Malaysia. *Higher Education*, *0123456789*. https://doi.org/10.1007/s10734-021-00689-5

Ball, K. F. (2012). *An investigation of the attitudes of school leaders toward the inclusion of students with disabilities in the general education setting*. The University of

Memphis.

Barber, J. P. (2012). Integration of Learning: A Grounded Theory Analysis of College Students' Learning. *American Educational Research Journal*, *49*(3), 590–617. https://doi.org/10.3102/0002831212437854

Barber, J. P. (2020). *Faciliting the Integration of Learning: Five Research-Based Practives to Help College Students Connect Learning Across Disciplines and Lived Experience*. Stylus Publishing, LLC.

Barnett, C., Miller, G., Polito, T., & Gibson, L. (2009). The Effect of an Integrated Course Cluster Learning Community on the Oral and Written Communication Skills and Technical Content Knowledge of Upper-Level College of Agriculture Students. *Journal of Agricultural Education*, *50*(2), 1–11. https://doi.org/10.5032/jae.2009.02001

Bell, P., Davis, E. A., & Linn, M. C. (1995). *The Knowledge Integration Environment:Theory and Design*. *October*, 14–21. https://doi.org/10.3115/222020.222043

Bicer, A., Capraro, R. M., & Capraro, M. M. (2013). Integrating Writing into Mathematics Classroom to Increase Students' Problem Solving Skills A R TIC LE I N F O A BS T RA C T. *International Online Journal of Educational Sciences*, *5*(2), 361–369. www.iojes.net

Bjorklund, S. A., Parente, J. M., & Sathianathan, D. (2004). Effects of faculty interaction and feedback on gains in student skills. *Journal of Engineering Education*, *93*(2), 153–160. https://doi.org/10.1002/j.2168-9830.2004.tb00799.x

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*(1), 29–51.

Bock, R. D. (1997). The nominal categories model. In *Handbook of modern item response theory* (pp. 33–49). Springer.

Boix Mansilla, V. (2010). *The Oxford handbook of interdisciplinarity*. Oxford: Oxford University Press.

Bonwell, C. C., & Eison, J. A. (1991). *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports.* ERIC.

Borrego, M., & Newswander, L. K. (2010). Definitions of interdisciplinary research: Toward graduate-level interdisciplinary learning outcomes. *Review of Higher Education, 34*(1), 61–84. https://doi.org/10.1353/rhe.2010.0006

Bosch, W. C., Hester, J. L., MacEntee, V. M., MacKenzie, J. A., Morey, T. M., Nichols, J. T., Pacitti, P. A., Shaffer, B. A., Tomascak, P. B., Weber, S. P., & Young, R. R. (2008). Beyond lip-service: An operational definition of "learning-centered college." *Innovative Higher Education, 33*(2), 83–98. https://doi.org/10.1007/s10755-008-9072-1

Brennan, R. L. (2001). An essay on the history and future of reliability from the perspective of replications. *Journal of Educational Measurement, 38*(4), 295–317.

Burton, K. (2006). Designing criterion-referenced assessment. *Journal of Learning Design, 1*(2), 73–82.

Burton, L. H. (2001). Curriculum: Retrospect and Prospect. *Music Educators Journal, 87*(5), 17–21.

Byrne, B. M. (2013). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Psychology Press.

Callaghan, N., & Bower, M. (2012). Learning through social networking sites - the critical role of the teacher. *Educational Media International*, *49*(1), 1–17. https://doi.org/10.1080/09523987.2012.662621

Casey, G., & Wells, M. (2015). Remixing to design learning: social media and peer-to-peer interaction. *Journal of Learning Design*, *8*(1). https://doi.org/10.5204/jld.v8i1.225

Castillo-Díaz, M., & Padilla, J.-L. (2013). How cognitive interviewing can provide validity evidence of the response processes to scale items. *Social Indicators Research*, *114*(3), 963–975.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative*. Thousand Oaks, CA: Sage.

Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Chubin, D. E., Porter, A. L., Rossini, F. A., & Connolly, T. (1986). *Interdisciplinary Research and Analysis*. MD: Lomond Publications.

Clark, A. M. (1998). The qualitative-quantitative debate: moving from positivism and confrontation to post-positivism and reconciliation. *Journal of Advanced Nursing*, *27*(6), 1242–1249.

Cohen, A. M., & Brawer, F. B. (2008). The American community college. *San*

Francisco, CA: Jossey-Bass, 374.

Collier, K. G. (1980). Peer-Group Learning in Higher Education: The development of higher order skills. *Studies in Higher Education*, *5*(1), 55–62. https://doi.org/10.1080/03075078012331377306

Colvin, J. W. (2007). Peer tutoring and social dynamics in higher education. *Mentoring and Tutoring: Partnership in Learning*, *15*(2), 165–181. https://doi.org/10.1080/13611260601086345

Corno, L., & Mandinach, E. B. (1983). The Role Of Cognitive Engagement in Classroom Learning and Motivation. *Educational Psychologist*, *18*(2), 88–108. https://doi.org/10.1080/00461528309529266

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. ERIC.

De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.

Dee Fink, L. (2009). Building an Integrated Student Learning Outcomes Assessment for General Education: Three Case Studies. *New Directions for Teaching and Learning*, *119*, 1–7. https://doi.org/10.1002/tl

Dillenbourg, P. (1999). *What do you mean by collaborative learning? What do you mean by "collaborative learning"?* https://telearn.archives-ouvertes.fr/hal-00190240

Dilly Fung. (2017). Outward-facing Student Assessments. In *Connected Curriculum for Higher Education* (p. 102). UCL Press. https://www.jstor.org/stable/j.ctt1qnw8nf.14

Dinan, S. E. (n.d.). The Importance of the First-Semester Experience: Learning Communities and Clustered Classes. *Breaking Barriers in Teaching and Learning*,

35.

Downing, S. M., & Haladyna, T. M. (1997). Test item development: Validity evidence from quality assurance procedures. *Applied Measurement in Education*, *10*(1), 61–82.

Edwards, A. (1996). *Interdisciplinary undergraduate programs: A directory*.

Ellis, B. B., & Mead, A. D. (2002). *Item analysis: Theory and practice using classical and modern test theory*.

Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. In *Item response theory for psychologists.* Lawrence Erlbaum Associates Publishers.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

Engelhard Jr, G. (2005). Guttman scaling. *Encyclopedia of Social Measurement*, 167–174.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. sage.

Glenn, D. M. (2005). *Developing a multisource feedback instrument: A comparison of CTT and IRT*. University of Houston.

Graebner, M. E., Martin, J. A., & Roundy, P. T. (2012). Qualitative data: Cooking without a recipe. *Strategic Organization*, *10*(3), 276–284.

Greene, D. B. (2003). Assessing General Education Learning Outcomes Achieved in the Inquiry-Guided, Self-Designed Major at North Carolina State University. *The Journal of General Education*, *52*(4), 304–316.

https://doi.org/10.1353/jge.2004.0013

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12*(3), 38–47.

Hartesveldt, C. Van, & Giordan, J. (2008). *Impact of transformative interdisciplinary research and graduate education on academic institutions. NSF 09-33*.

Hartup, W. W. (1989). Social relationships and their developmental significance. *American Psychologist*, *44*(2), 120.

Haynes, S. N., Richard, D., & Kubany., E. S. (1995). Validity, Content Methods, psychological assessment: A functional approach to concepts and. *Psychological Assessment*, *7*(3), 238.

Hegji, A. (2017). An Overview of Accreditation of Higher Education in the United States. CRS Report R43826, Version 7. Updated. *Congressional Research Service*.

Holley, K. (2017). Administering interdisciplinary programs. In *The Oxford handbook of interdisciplinarity*.

Hoxby, C. M., Winston, G. C., & Zimmerman, D. J. (2013). Peer Effects in Higher Education. In *College Choices: Vol. I* (Issue September). https://doi.org/10.7208/chicago/9780226355375.003.0010

Hursh, B., Haas, P., & Moore, M. (1983). An Interdisciplinary Model to Implement General Education. *The Journal of Higher Education*, *54*(1), 42. https://doi.org/10.2307/1981644

Hurtado, S. (2007). Linking diversity with the educational and civic missions of higher

education. *Review of Higher Education*, *30*(2), 185–196. https://doi.org/10.1353/rhe.2006.0070

Integrated Postsecondary Education Data System. (2021). *About IPEDS*. National Center for Education Statistics. https://nces.ed.gov/ipeds/about-ipeds

J. Fredericks Volkwein. (2009). The Assessment Context: Accreditation, Accountability, and Performance. *New Directions for Institutional Research*, *14*(7), 1–11. https://doi.org/10.1002/ir

Jacobs, H. H. (1989). Interdisciplinary Curriculum: design and implematation. In *Ìkala, Revista De Lenguaje Y Cultura* (Vol. 8, Issue Curriculum design).

Jacobs, H. H., & Borland, J. H. (1986). The interdisciplinary concept model: Theory and practice. *Gifted Child Quarterly*, *30*(4), 159–163.

Jamieson, S. (2004). Likert scales: How to (ab) use them? *Medical Education*, *38*(12), 1217–1218.

Johnson, D. K., Ratcliff, J. L., & Gaff, J. G. (2004). A Decade of Change in General Education. *New Directions for Higher Education*, *2004*(125), 9–28. https://doi.org/10.1002/he.136

Jones, C. (2009). Interdisciplinary Approach: Advantages, Disadvantages, and the Future Benefits of Interdisciplinary Studies. *Essai*, *7*(26). http://dc.cod.edu/essai/vol7/iss1/26

Joreskog, K. G., & Sorbom, D. (1984). LISREL VI: Analysis of linear structural relationships by the method of maximum likelihood. *Chicago: National Educational Resources*.

Kane, M. (2010). Validity and fairness. *Language Testing*, *27*(2), 177–182.

Kim, K., Sharma, P., Land, S. M., & Furlong, K. P. (2013). Effects of Active Learning on Enhancing Student Critical Thinking in an Undergraduate General Science Course. *Innovative Higher Education*, *38*(3), 223–235. https://doi.org/10.1007/s10755-012-9236-x

King, Alison. (1993). From sage on the stage to guide on the side. *College Teaching*, *41*(1), 30–35.

King, AR, & Brownell, J. (1966). *The curriculum and the disciplines of knowledge: A theory of curriculum practice*.

King, P. M., Brown, M. K., Lindsay, N. K., & Vanhecke, J. R. (2007). *Liberal Arts Student Learning Outcomes : An Integrated Approach*. *October*, 2–9. https://doi.org/10.1002/abc.2

Klein, J. T., & Newell, W. H. (1997). *Advancing interdisciplinary studies. Handbook of the undergraduate curriculum: A comprehensive guide to purposes, structures, practices, and change*. San Francisco, CA: Jossey- Bass.

Klein, J. T. (1990). Interdisciplinarity: History, theory, and practice. *Wayne State University Press*.

Klein, J. T. (1999). *Mapping Interdisciplinag Studies*.

Klein, J. T. (2005). Integrative Learning and Interdisciplinary Studies. *Peer Review*, *7*(4), 8–10. https://doi.org/http://dx.doi.org/10.1108/17506200710779521

Klein, J, & Newell, W. (1996). *Interdisciplinary studies. Handbook on the Undergraduate Curriculum*. Jossey-Bass.

Klein, JT. (2005). *The Changing American Academy: Humanities, Culture, and Interdisciplinarity*.

Koch, W. R. (1983). Likert scaling using the graded response latent trait model. *Applied Psychological Measurement, 7*(1), 15–32.

Kuh, G. D., Jankowski, N., Ikenberry, S. O., & Kinzie, J. (2014). Knowing What Students Know and Can Do: The Current State of Student Learning Outcomes Assessment in U . S . Colleges and Universities. ... *Outcomes ...*, *January*, 1–19. www.learningoutcomesassessment.org

LaPointe, D. K., & Gunawardena, C. N. (2004). Developing, testing and refining of a model to understand the relationship between peer interaction and learning outcomes in computer-mediated conferencing. *International Journal of Phytoremediation*, *25*(1), 83–106. https://doi.org/10.1080/0158791042000212477

Lattuca, L. R. (2001). Creating Interdisciplinarity: Interdisciplinary Research and Teaching among College and University Faculty. In *The Journal of Higher Education* (Vol. 73, Issue 6). Vanderbilt University Press. https://doi.org/10.1080/00221546.2002.11777178

Lattuca, L. R., Knight, D., & Bergom, I. (2013). Developing a measure of interdisciplinary competence. *The International Journal of Engineering Education*, *29*(3), 726–739.

Lattuca, L. R., Voight, L. J., & Fath, K. Q. (2004). Does interdisciplinarity promote learning? Theoretical support and researchable questions. *The Review of Higher Education*, *28*(1), 23–48. https://muse.jhu.edu/article/172217/summary

Lie, A. (2002). *Cooperative Learning (Cover Baru)*. Grasindo.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*.

Linn, M. C. (1995). Designing computer learning environments for engineering and computer science: The scaffolded knowledge integration framework. *Journal of Science Education and Technology*, *4*(2), 103–126.

Linn, M. C. (2000). Designing the knowledge integration environment. *International Journal of Science Education*, *22*(8), 781–796. https://doi.org/10.1080/095006900412275

Lipson, M. Y., Valencia, S. W., Wixson, K. K., Peters, C. W., Lipson, M. A. R. J. Y., Valencia, S. W., Wixson, K. K., & Peters, C. W. (1993). Integration and Thematic Teaching : Integration to Improve Teaching and Learning. *National Council of Teachers of English*, *70*(4), 252–263.

Lowenstein, M. (2015). General education, advising, and integrative learning. *The Journal of General Education*, *64*(2), 117–130.

Mansilla, V. B. (2005). Assessing Student Work at Disciplinary Crossroads. *Change: The Magazine of Higher Learning*, *37*(1), 14–21. https://doi.org/10.3200/chng.37.1.14-21

Mansilla, V. B., & Duraising, E. D. (2007). Targeted assessment of students' interdisciplinary work: An empirically grounded framework proposed. *Journal of Higher Education*. https://doi.org/10.1353/jhe.2007.0008

Marie-France Orillion. (2009). Interdisciplinary Curriculum and Student Outcomes: The Case of a General Education Course at a Research University. *The Journal of*

General Education, *58*(1), 1–18. https://doi.org/10.1353/jge.0.0032

Marshall, D., Beck, S., Coghlan, L., & Kimbro, E. (2005). Understanding Student Learning Outcomes through Narrative Transcript Analysis: Assessing General Education in an Institution without Grades or Required Courses. *Assessment Update - Progress, Trends, and Practices in Higher Education*, *17*(5), 4.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174.

Masters, G. N. (2016). Partial credit model. In *Handbook of Item Response Theory, Volume One* (pp. 137–154). Chapman and Hall/CRC.

McMahon, W. W. (2009). Higher learning, greater good: The private and social benefits of higher education. In *Higher Learning, Greater Good: The Private and Social Benefits of Higher Education*. https://doi.org/10.1080/13603101003780376

McQuillan, M., Purves, R., & Macdonald, G. (2022). *Post-Theory*. Edinburgh University Press.

Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research*, *45*(1), 35–44.

Miller, R. (2000). Creating learning communities: Models, resources, and new ways of thinking about teaching and learning. *Foundation for Educational Renewal*.

Miller, Ross. (2005). Integrative learning and assessment. *Peer Review*, *7*(4), 11–14.

Millsap, R. E., Van der Ark, L. A., Bolt, D. M., & Woods, C. M. (2016). *New Developments in Quantitative Psychology*. Springer.

Moran, J. (2010). *Interdisciplinarity*. Routledge.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *ETS Research Report Series, 1992*(1), i–30.

Muraki, E. (1997). A generalized partial credit model. In *Handbook of modern item response theory* (pp. 153–164). Springer.

Nassar-McMillan, S. C., & Borders, L. D. (2002). Use of focus groups in survey item development. *The Qualitative Report*, *7*(1), 1–12.

Nelson, T. F., Niskodé-dossett, A. S., Kuh, G. D., Laird, T. F. N., & Niskod, A. S. (2009). *Linked references are available on JSTOR for this article : What General Education Courses Contribute to Essential Learning Outcomes*. *58*(2), 65–84.

Newell, W. H. (2001). A Theory of Interdisciplinary Studies. *Issues in Integrative Studies*, *25*(19), 1–25.

Nowacek, R. S. (2005). A Discourse-Based Thoery Of Interdisplinary Connections. *The Journal of General Education*, *54*(3), 171–195.

Nunnaly, J. C., & Bernstein, I. H. (1994). Construction of conventional tests. *Psychometric Theory*, 293–323.

Omplication, A., Ost, P., Krämer, S., Broschewitz, J., Kirsten, H., Sell, C., Eichfeld, U., Struck, M. F., Либерти, C., Ao, C., Blanch, A., Costescu, F., Slinger, P., Ezalia, E., R, I. E., Elizabeth, G., My, W. A. N. H., Norhanim, A., Wahidah, A., … Caffaratti, E. (2006). No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title. *Irtualm-VBiblioteca Ep*, *March*, 1–92. https://doi.org/10.1590/s1809-98232013000400007

Pappas, C. C., Kiefer, B. Z., & Levstik, L. S. (1990). *Ar integrated language perspective*

in the elementary school: Theory into action. New York: Longman.

Pascarella, E., & Terenzini, P. (2005). *How College Affects Students: A Third Decade of Research*. Jossey-Bass. https://eric.ed.gov/?id=ED498537

Ploetzner, R., Dillenbourg, P., Preier, M., & Traum, D. (1999). Learning by Explaining to Oneself and to Others. In *people.ict.usc.edu*. Pergamon Press. https://people.ict.usc.edu/traum/public_html/Papers/esf-ploztner.pdf

Polin, L. (2004). Learning in Dialogue With a Practicing Community. *Learner-Centered Theory and Practice in Distance Education: Cases from Higher Education*. https://www.academia.edu/download/3451738/Learning_in_dialogue_with_a_practicing_community.pdf#page=40

Purzer, S., & Cardella, M. E. (n.d.). Instrument Development Model: A Map based on Messick's Unified Theory of Validity. *Licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License*.

Renn, K. A., & Arnold, K. D. (2003). Reconceptualizing research on college student peer culture. *Journal of Higher Education*, *74*(3), 261–291. https://doi.org/10.1080/00221546.2003.11780847

Repko, A. (2008). *Interdisciplinary research: Theory and methods*.

Richman, W. A., & Ariovich, L. (2013). All-in-one: Combining grading, course, program, and general education outcomes assessment. *Occasional Paper, 19*.

Roche, M. W. (2010). *Why choose the liberal arts?* University of Notre Dame Pess.

Rosenzweig, K. (2009). *Are Today's General Education Teachers Prepared to Meet the Needs of Their Inclusive Students?*

Roszkowski, M. (2003). *the Nuts and Bolt S of. 2011*, 169–188.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461–464.

Seifert, T. A., Goodman, K. M., Lindsay, N., Jorgensen, J. D., Wolniak, G. C., Pascarella, E. T., & Blaich, C. (2008). The effects of liberal arts experiences on liberal arts outcomes. *Research in Higher Education*, *49*(2), 107–125. https://doi.org/10.1007/s11162-007-9070-7

Smith, B. L., MacGregor, J., Matthews, R., & Gabelnick, F. (2009). *Learning communities: Reforming undergraduate education*.

Smith, K. (1979). Learning together and alone: cooperation, competition, and individualization. *NACTA Journal*, *23*(3), 23–26.

Snow, S. G. . (1973). Correlates of Faculty-Student Interaction. *Sociology of Education*, *46*(4), 489–498. https://www.jstor.org/stable/2111902

Songer, N. B., & Linn, M. C. (1991a). How do students' views of science influence knowledge integration? *Journal of Research in Science Teaching*, *28*(9), 761–784. https://doi.org/10.1002/tea.3660280905

Songer, N. B., & Linn, M. C. (1991b). How do students' views of science influence knowledge integration? *Journal of Research in Science Teaching*, *28*(9), 761–784.

Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. *The American Journal of Psychology*, 161–169.

Spearman, C. (1913). Correlations of sums or differences. *British Journal of Psychology*, *5*(4), 417.

Stark, J. S., & Lattuca, L. R. (1997). *Shaping the college curriculum: Academic plans in action.*

Stassen, M. L. A. (2019). *Assessing Students' Integrative Learning Across Curricular and Co-Curricular Experiences*. https://www.umass.edu/oapa/program-assessment/general-education-assessment/integrative-experience-ie/assessment/assessing

Steedle, J. T. (2012). Selecting value-added models for postsecondary institutional assessment. *Assessment and Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2011.560720

Stefani, L. (2008). Assessment in interdisciplinary and interprofessional programs: Shifting paradigms. In *Interdisciplinary Learning and Teaching in Higher Education* (pp. 44–57). Routledge.

Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics, 6th Edn. Northridge. *CA: California State University*.

Tondreau, D. M. (2003). *Weaving storylines in adult education: The narrative construction of self, society, and structures of knowledge*. Union Institute and University.

Topping, K., & Ehly, S. (1998). *Peer-assisted learning*. Routledge. https://books.google.com/books?hl=en&lr=&id=pDSRAgAAQBAJ&oi=fnd&pg=PP1&dq=Peer-

assisted+learning&ots=sKwll64oNb&sig=YJXczSDap33GaTS4m_MOhsMEUNc

Traub, R. E. (1997). Classical test theory in historical perspective. *Educational Measurement, 16*, 8–13.

van den Berg, I., Admiraal, W., & Pilot, A. (2006). Designing student peer assessment in higher education: Analysis of written and oral peer feedback. *Teaching in Higher Education, 11*(2), 135–147. https://doi.org/10.1080/13562510500527685

Virginia tech. (n.d.). *Facts About Virginia Tech*. https://www.vt.edu/about/facts-about-virginia-tech.html

Walker, D., & Myrick, F. (2006). Grounded theory: An exploration of process and procedure. *Qualitative Health Research, 16*(4), 547–559.

Wang, Z., & Osterlind, S. J. (2013). Classical test theory. In *Handbook of quantitative methods for educational research* (pp. 31–44). Springer.

William H. Newell. (1990). Interdisciplinary Curriculum Development. *Issues in Integrative Studies*, *8*, 69–86. https://doi.org/10.2307/3192194

Wilson, L. R. (1946). General Education in a Free Society; Report of the Harvard Committee (Book Review). In *College & Research Libraries* (Vol. 7, Issue 2). Harvard University Press. https://doi.org/10.5860/crl_07_02_186

Wolf-Wendel, L., Ward, K., & Kinzie, J. (2009). A Tangled Web of Terms: The Overlap and Unique Contribution of Involvement, Engagement, and Integration to Understanding College Student Success. *Journal of College Student Development*, *50*(4), 407–428. https://doi.org/10.1353/csd.0.0077

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size

requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, *73*(6), 913–934.

Yang, M. (2009). Making interdisciplinary subjects relevant to students: An interdisciplinary approach. *Teaching in Higher Education, 14*(6), 597–606.

Yen, W. M., Fitzpatrick, A. R., & Brennan, R. L. (2006). *Educational measurement*. Westport, CT: Praeger Publishers.

Youatt, J., & Wilcox, K. A. (2008). Intentional and integrated learning in a new cognitive age: A signature pedagogy for undergraduate education in the twenty-first century. *Peer Review*, *10*(4), 24.

Youngblood, D. (2007). Multidisciplinarity, interdisciplinarity, and bridging disciplines: A matter of process. *Journal of Research Practice*, *3*(2), M18–M18. http://jrp.icaap.org/index.php/jrp/article/view/104

Yu, L., Shek, D. T. L., & Zhu, X. (2019). General Education Learning Outcomes and Demographic Correlates in University Students in Hong Kong. *Applied Research in Quality of Life*, *14*(5), 1165–1182. https://doi.org/10.1007/s11482-018-9645-9

Volkwein, J. F. (2007). Assessing institutional effectiveness and connecting the pieces of a fragmented university. Fixing the fragmented university, 145-180.

Appendix A

IRB Approval

Division of Scholarly Integrity and

Research Compliance

Institutional Review Board

North End Center, Suite 4120 (MC 0497)

300 Turner Street NW

Blacksburg, Virginia 24061

540/231-3732

irb@vt.edu

http://www.research.vt.edu/sirc/hrpp

MEMORANDUM

DATE: February 10, 2020

TO: Stephen Michael Biscotte, Mengyun Li

FROM: Virginia Tech Institutional Review Board (FWA00000572, expires October 29,

2024)

PROTOCOL TITLE: Pathways Minor Program Evaluation

IRB NUMBER: 19-1129

Based on the submitted project description and items listed in the Special Instructions section found on Page 2, the Virginia Tech IRB has determined that the

proposed activity is not research involving human subjects as defined by HHS and FDA regulations.

Further review and approval by the Virginia Tech HRPP is not required because this is not human research. This determination applies only to the activities described in the submitted project description and does not apply should any changes be made. If changes are made you must immediately submit an Amendment to the HRPP for a new determination. Your amendment must include a description of the changes and you must upload all revised documents. At that time, the HRPP will review the submission activities to confirm the original "Not Human Subjects Research" decision or to advise if a new application must be made.

If there are additional undisclosed components that you feel merit a change in this initial determination, please contact our office for a consultation. Please be aware that receiving a "Not Human Subjects Research" Determination is not the same as IRB review and approval of the activity. You are NOT to use IRB consent forms or templates for these activities. If you have any questions, please contact the Virginia Tech HRPP office at 540-231-3732 or irb@vt.edu.

PROTOCOL INFORMATION:

Determined As: Not Human Subjects Research

Protocol Determination Date: **February 10, 2020**

Appendix B

Invitation Letter I

Hi Dr. **,

I am Mengyun from the Office of Undergraduate Education and I work with Dr. *** to support the Pathways Minor General Education program.

As you know, the Pathways Minors are a very important part of that program. We designed a survey to gather student's perspectives on their high-level learning outcomes and experiences associated with Pathways Minors. This will give us a sense of whether the minors are meeting the goals of the program and help us together make improvements to the courses and minor in the future. The target group is the minor students in the capstone course since they would have had the most experiences in this minor. We will provide you with the results to help you identify successes and make improvements as you see fit.

If willing, we will get back to you at the beginning of April to distribute the survey to your students by the end of the semester. We recommend posting the link to Canvas as a credit-bearing assignment but are open to whatever works for you to ensure a high response rate.

Best,

Mengyun

Appendix C

Invitation Letter II

Hi Dr.*,

I am Mengyun who is working with Dr. * to support the Pathways General Education program.

As we discussed in January, we plan to start the process of evaluation for the minor. We designed a survey to gather student's perspectives on their high-level learning outcomes and experiences associated with Pathways Minors. This will give us a sense of whether the minors are meeting the goals of the program and help us together make improvements to the courses and minor in the future. The target group is the minor students in the capstone course since they would have had the most experiences in this minor.

We will provide you with the results to help you identify successes and make improvements as you see fit. Please distribute the survey to your students by the end of the semester, we really appreciate it. We will provide you the list of student names before the final exam if you choose the credit-bearing assignment this semester.

(Link)

Thank you very much for your time and patience.

Mengyun

Appendix D

Instrument in Spring 2021

THEME 1: I feel the courses in the minor were connected by a common theme.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

THEME 2: I feel the number of classes I completed for this minor is sufficient to meet the goals of the minor.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

THEME 3: The courses in this minor offered multiple perspectives on the topics and issues central to this minor.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

DISGR1: The learning goals of this minor are clear to me.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

DISGR2: I feel I have learned the fundamental concepts related to this minor.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

DISGR3: I recognize the commonalities between fields or disciplines included in this minor

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

DISGR4: I recognize the differences between fields or disciplines included in this minor.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

KNOWLEDGE1: I can explain things I learned from this minor to others.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

KNOWLEDGE2: In assignments, I have applied knowledge across different fields or disciplines central to this minor.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

KNOWLEDGE3: I have expanded my own views of the concepts and issues central to this minor.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

KNOWLEDGE4: I have changed the way I view concepts and issues central to this minor.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

KNOWLEDGE5_text: Please give one or two examples of times you made connections between courses in this minor.

BEYMIN1: I have identified problems or issues related to this minor in the real world.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

BEYMIN2: I have used the knowledge gained from this minor to solve problems in my everyday life.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

BEYMIN3: I have made connections between things I learned in this minor and things I already knew.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

BEYMIN4: I have made connections between things I learned in this minor and my major.(Deleted in Spring 2022)

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

BEYMIN5: I have made connections between things I learned in this minor to my future career plan.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

BEYMIN6: Please give one or two examples of times when you connected experiences in this minor with your experiences in the real world.

LE1: Throughout the minor, I feel the classroom environment (communication, support, available help) has supported my learning.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

LE2: Throughout the minor, I feel instructors have helped me solve problems.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

LE3: Over the course of the minor, how many times did you interact with the instructor outside of class?

1 Many

2 Some

4 A few

5 None

LE4: Over the course of the minor, how many times did the instructors create time for you to think about the content taught in the class?

1 Many

2 Some

4 A few

5 None

LE5: Over the course of the minor, how many times were you asked to reflect on, revise, and resubmit an assignment?

1 Many

2 Some

4 A few

5 None

LE6: Over the course of the minor, how many times did you discuss your thoughts related to the course and minor topics with classmates?

1 Many

2 Some

4 A few

5 None

LE7: Over the course of the minor, how many times did you work collaboratively?

(Deleted in 2022 Spring)

1 Many

2 Some

4 A few

5 None

LE8: Over the course of the minor, in how many group projects did you participate with students from majors different than yours?

1 Many

2 Some

4 A few

5 None

LE9: During the course of the minor, how many times did you interact with other students enrolled in the minor outside of class?

1 Many

2 Some

4 A few

5 None

LE10_text: Please give one or two examples how you interact with the people (faculty, students) in this minor.

Appendix E

Instrument in Spring 2022

THEME 1: I feel the courses in the minor were connected by a common theme.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

THEME 2: I feel the number of classes I completed for this minor is sufficient to meet the goals of the minor.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

THEME 3: The courses in this minor offered multiple perspectives on the topics and issues central to this minor.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

THEME 4: The classes I took early on in the minor prepared me well for the classes I took later on in the minor. (Added in 2022 Spring)

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

DISGR1: The learning goals of this minor are clear to me.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

DISGR2: I feel I have learned the fundamental concepts related to this minor.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

DISGR3: I recognize the commonalities between fields or disciplines included in this minor.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

DISGR4: I recognize the differences between fields or disciplines included in this minor.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

KNOWLEDGE1: I can explain things I learned from this minor to others.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

KNOWLEDGE2: In assignments, I have applied knowledge across different fields or disciplines central to this minor.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

KNOWLEDGE3: I have expanded my own views of the concepts and issues central to this minor.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

KNOWLEDGE4: I have changed the way I view concepts and issues central to this minor.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

KNOWLEDGE5: I have reflected on the knowledge I have learned throughout this minor.(Added in Spring 2022)

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

BEYMIN1: I have identified problems or issues related to this minor in the real world.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

BEYMIN2: I have used the knowledge gained from this minor to solve problems in my everyday life.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

BEYMIN3: I have made connections between things I learned in this minor and things I already knew.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

BEYMIN5: I have made connections between things I learned in this minor to my future career plan.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

BEYMIN6: My experiences in this minor have inspired me to think about applying my studies in new, innovative ways.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

LE1: Throughout the minor, I feel the classroom environment (communication, support, available help) has supported my learning.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

LE2: Throughout the minor, I feel instructors have helped me solve problems.

1 Strongly agree

2 Agree

3 Somewhat agree

4 Somewhat disagree

5 Disagree

6 Strongly disagree

LE3: Over the course of the minor, how many times did you interact with the instructors outside of class?

1 Many

2 Some

4 A few

5 None

LE5: Over the course of the minor, how many times did the instructor ask you to reflect on, revise, or resubmit an assignment?

1 Many

2 Some

4 A few

5 None

LE6: Over the course of the minor, how many times did you discuss your thoughts related to the course and minor topics with classmates?

1 Many

2 Some

4 A few

5 None

LE4: Over the course of the minor, how many times did the instructors create time for you to think about the content taught in the class?

1 Many

2 Some

4 A few

5 None

LE8: Over the course of the minor, in how many group projects did you participate with students from majors different than yours?

1 Many

2 Some

4 A few

5 None

LE9: During the course of the minor, how many times did you interact with other students enrolled in the minor outside of class?

1 Many

2 Some

4 A few

5 None

Appendix F

Sample Report of the Survey

Sample Report of the Survey, Part 1 (n=)*

Question	Strongly Agree	Agree	Somewhat Agree	Somewhat Disagree	Disagree	Strongly Disagree	Weighted Mean
I feel the courses in the minor were connected by a common theme.							
I feel the number of classes I completed for this minor is sufficient to meet the goals of the minor.							
The courses in this minor offered multiple perspectives on the topics and issues central to this minor.							
The classes I took early on in the minor prepared me well for the classes I took later on in the minor.							
The learning goals of this minor are clear to me.							

I feel I have learned the fundamental concepts related to this minor.							
I recognize the commonalities between fields or disciplines included in this minor.							
I recognize the differences between fields or disciplines included in this minor.							
I can explain things I learned from this minor to others.							
In assignments, I have applied knowledge across different fields or disciplines central to this minor.							
I have expanded my own views of the concepts and issues central to this minor.							
I have changed the way I view concepts and							

issues central to this minor.							
I have reflected on the knowledge I have learned throughout this minor.							
I have identified problems or issues related to this minor in the real world.							
I have used the knowledge gained from this minor to solve problems in my everyday life.							
I have made connections between things I learned in this minor and things I already knew.							
I have made connections between things I learned in this minor to my future career plan.							
My experiences in this minor have inspired me to think about applying my							

studies in new, innovative ways.							
Throughout the minor, I feel the classroom environment (communication, support, available help) has supported my learning.							
Throughout the minor, I feel instructors have helped me solve problems.							

This chart contains the distribution of responses for questions where respondents chose from the six choices listed. The weighted mean is shown in the final column to the right. A weighted mean of 6.000 would indicate all respondents selected "strongly agree."

Sample Report of the Survey, Part 2 (n=)*

Question	Many	Some	A few	None	Weighted Mean
Over the course of the minor, how many times did you interact with the instructors outside of class?					
Over the course of the minor, how many times did the instructor ask you to reflect on, revise, or resubmit an assignment?					
Over the course of the minor, how many times did you discuss your thoughts related to the course and minor topics with classmates?					

Over the course of the minor, how many times did the instructors create time for you to think about the content taught in the class?					
Over the course of the minor, in how many group projects did you participate with students from majors different than yours?					
During the course of the minor, how many times did you interact with other students enrolled in the minor outside of class?					

This chart contains the distribution of responses for questions where respondents chose from the four choices listed. The weighted mean is shown in the final column to the right. A weighted mean of 4.000 would indicate all respondents selected "many."

Respondents were able to choose more than one response to this question.

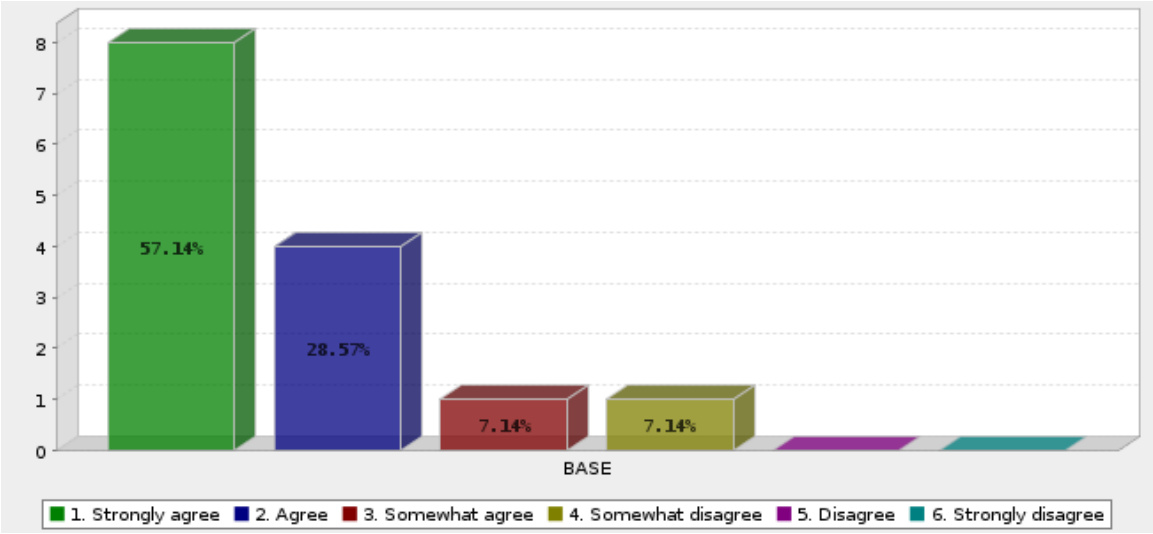
Sample Report of the Survey, Part 3

Question	Advisor	Friend	Instructor in a class	Website	Social Media	A flier or poster	Parent or guardian	Other
How did you hear about this minor?								

Question: What are 2-3 of the most positive takeaways from completing this minor?
Responses:
Question: What are 2-3 ways you think the minor could be improved?
Responses:
Question: Why did you choose to enroll in this minor?
Responses:

Appendix G

Raw Question Sample Report



	Answer	Count	Percent
1.	Strongly agree	8	57.14%
2.	Agree	4	28.57%
3.	Somewhat agree	1	7.14%
4.	Somewhat disagree	1	7.14%
5.	Disagree	0	0.00%
6.	Strongly disagree	0	0.00%
	Total	14	100%

Mean: 1.643	Confidence Interval @ 95%: [1.156 - 2.129]	Standard Deviation: 0.929	Standard Error: 0.248