Summer 7-22-2021

Threshold concepts in web development: the impact of education and experience on the perceptions of practitioners

Michael Lee Mick
DePaul University, MMICK@depaul.edu

THRESHOLD CONCEPTS IN WEB DEVELOPMENT: THE IMPACT OF EDUCATION AND EXPERIENCE ON THE PERCEPTIONS OF PRACTITIONERS

BY

MICHAEL LEE MICK

A DISSERTATION SUBMITTED TO THE SCHOOL OF COMPUTING, COLLEGE OF

COMPUTING AND DIGITAL MEDIA OF DEPAUL UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPAUL UNIVERSITY

CHICAGO, ILLINOIS

2021

DePaul University

College of Computing and Digital Media

Dissertation Verification

This doctoral dissertation has been read and approved by the dissertation committee below according to the requirements of the Computer and Information Systems PhD program and DePaul University.

Name: Michael Lee Mick

Title of dissertation:

Threshold Concepts in Web Development: The Impact of Education and Experience on the Perceptions of Practitioners

Date of Dissertation Defense:

July 22, 2021

Craig Miller, PhD
Dissertation Co-Advisor*

Serdar Turedi, PhD
Dissertation Co-Advisor

Amber Settle, PhD
1st Reader

Theresa Steinbach, PhD
2nd Reader

** A copy of this form has been signed, but may only be viewed after submission and approval of FERPA request letter.*

THRESHOLD CONCEPTS IN WEB DEVELOPMENT: THE IMPACT OF EDUCATION AND EXPERIENCE ON THE PERCEPTIONS OF PRACTITIONERS

ABSTRACT

The Threshold Concepts (TCs) framework posits characteristics of critical concepts that enable a learner to see things in a new and unforgettable way that allows the incorporation of cognitive approaches and skills needed by professionals in the discipline. All previous research has been qualitatively oriented toward discovery of candidate TCs, without, so far, quantitative studies of the candidate TCs recognition within the professional ranks. An underrepresented group in previous research is that of Web development professionals.

This study describes initial quantitative research utilizing the TCs framework in the Web development area to measure the perception of candidate TCs by professionals in the field. This study develops and utilizes a new survey instrument to determine how a national sample of current Web development professionals perceive a candidate TC and how their work experience and level of education impact their perceptions. The particular candidate TC, Separation of Content, Presentation, and Behavior (SCPB), and its appropriate identifying characteristics are selected from results of previous qualitative research.

This study first does an exploratory analysis using SPSS on results from an initial group surveyed and then uses results from a later group for a confirmatory analysis with PLS-SEM. The exploratory analysis reduces the contributing factors used to identify the TC, and these factors and their measures are then used with the latter group in the confirmatory analysis. The factors identified for use were the characteristics of Transformative, Troublesome, Irreversible, and Bounded. The PLS-SEM analysis confirms that SCPB is perceived as a TC by working Web development professionals by virtue of it containing these characteristics, as the theory indicates.

The study looks at work experience, in terms of years in computing and years specifically in Web development, and education, in terms of education level (degree(s) received, if any), degree in a computing field, and years since last in formal education. All of these indicate a significant influence on perception, either positively or negatively, except for the last item which does not exhibit statistical significance in this case.

DEDICATION

For their support, love, encouragement, and prayers, I dedicate this work to my family and many friends. Although space does not allow listing all, I want to express special gratitude to family. My wife, Sandy, is an encourager by nature and never doubted the final outcome, even when I was uncertain. Along with our daughters, Heather and Karin, they have borne countless evenings when I was at class or working at home. I trust that the times I did set aside for you balanced your sacrifice. To our extended family, church friends, choir family, and small Bible study group, your prayers and thoughtful encouragement meant more than simply saying thanks can convey. You share in this accomplishment with us.

I also dedicate this to parents, my mother, Ruth, my in-laws, Ed and Adena, and my birth mother, Marla. My only regret is that you are no longer here with us at the completion.

My dedications would be incomplete without mentioning Gene and Phyllis Borman. Your confidence in me, even beyond material help, provided continued motivation for me to stay at the task until completed.

Most of all, I dedicate this dissertation to the One who made it all possible. Lord Jesus, You provided everything I needed, when needed, to see this through, and taught me so many life lessons along the way. May You receive due praise through my work and all of my life.

Soli Deo Gloria

ACKNOWLEDGEMENTS

No research work stands alone. It is a symphony of contributions without any one of which the work would never reach fruition. I would like to express my thanks and profound gratitude to those whose efforts directly impacted this writing.

First, to the members of my committee, who each provided invaluable and unique contributions to the completion of this work from its inception through final draft. I have learned much from you. It certainly is no overstatement, and quite possibly an understatement, to say that without your assistance, this work would not exist. You provided the inspiration, guidance, and oversight to the best type of dissertation: one which is (finally!) done.

Also, my thanks to the school librarians who helped guide me to references in times of need. No longer only masters of the stacks, but of digital realms as well! You are an unsung and yet essential part in most, if not all, scholarly endeavors. Such works as this one depend upon your benevolent and understanding assistance, often when the need is urgent.

And to John Glatz, director of advising, who has helped with my questions throughout the years. You showed great patience, encouragement, and practical assistance in helping me navigate administrative systems so that I could continue with my efforts unrestricted.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Web development as a Discipline

The influence of the Web is both ubiquitous and ongoing. The Web affects many sectors of life as use of the Web becomes central in diverse areas, such as business, government, education, social networks, and healthcare organizations (Gibson, 2006). Since the beginning of Web 2.0 with the move from static to dynamic web sites, the development of such sites also has grown increasingly complex (Murugesan & Deshpande, 2002). In general, learning to program is difficult and the rapid changes in the Web only add to the difficulty for new learners of software development for the Web (Aghaei, Nematbakhsh, & Farsani, 2012; Bennedsen & Caspersen, 2007; Bosse & Gerosa, 2016; Russo, 2000). There are numerous reasons why software development for the Web particularly is difficult. Some general areas are:

1) Users. The user base, at least for publicly accessible sites, if not for private intranet sites, is very different from that for traditional applications. This public user base is very large, operating in diverse environments (Chen, Chen, Miao, & Wang, 2013; Russo, 2000). This makes traditional user-centered requirements gathering impossible, beyond a representative sample. In-person end user training for such (publicly accessible) Web-based applications is impossible, requiring sites to be intuitively navigable. Users who access a given site for legitimate reasons may take unexpected actions, including leaving the site suddenly and may rarely, or never, return. Given the public nature of Web sites, the size of the user base can vary greatly at any point in time, requiring the design for scalability as being critical, at least for publicly accessible sites (Cardellini, Colajanni, & Yu, 1999; Dias, Kish, Mukherjee, & Tewari, 1996; Gray, 2001; O'Reilly, 2007).

2) Security. The users are unpredictable in purpose, as any given user may be of innocent or nefarious intent. Those trying to access a given site may not always be authorized users, or even human at all. The ability to secure such a site against unauthorized access and against automated access or attack becomes more complicated even as more applications move to a Web context, and the need to secure sites is of increasing concern (Miller & Connolly, 2015; Patrick & Fields, 2017).
3) Devices. Numerous factors affect quality of service of a Web-based application (Murugesan, 2008). Although the typical user interface is a Web browser, users bring a variety of devices to bear. These devices vary in available memory, display characteristics, and even in the specific Web browser and browser software version utilized. Further, the configuration of said browser and its capabilities, such as the ability to store cookies or run scripts, may be necessary for the application to function properly and yet are not within the control of the developer. The bandwidth available to the user via a given device can vary greatly, from device to device and over time and location, thus affecting application responsiveness and user satisfaction. Again, these variables are outside of the control of the developer, adding to the challenge of the development task (Le Pallec, Marvie, Rouillard, & Tarby, 2010).
4) Change. Compounding these difficulties is the relatively rapid change of technologies used to create such applications, to say nothing of the increasing time compression in the development of software for use on the Internet (Baskerville, Ramesh, Levine, Pries Heje, & Slaughter, 2003; Hoar & Connolly, 2016). Indeed, just keeping up with changing technology is challenging enough for instructors, not to mention their students (Connolly, 2019; Miller & Dettori, 2008).

Thus, the discipline of Web development is one of increasing change and diversity, even in a field like computing, where a certain degree of such is the norm. These factors place a considerable burden on those who teach, not only to stay current but also to decide what topics and concepts are most important to include in time-constrained quarter- or semester-long courses on Web development (Miller & Connolly, 2015).

1.2 Motivation for the study

Students enroll in university ostensibly to learn. Potential employers and/or graduate schools (depending on the post-graduation goals of the student) expect students to have learned certain concepts such that they are sufficiently prepared to enter their respective disciplines. Students enrolling in Web development courses face a difficult learning curve, given the challenges listed above.

While students will differ in their success at mastering required concepts, instructors presumably intend for their students to be able to think and practice as professionals in the field. University educators have engaged in research into the processes of teaching and learning in various disciplines for many decades, as evidenced by the founding of the ACM Special Interest Group for Computer Science Education (SIGCSE) in 1968, with multiple conferences applicable for Scholarship of Teaching and Learning (SoTL) publications. Other long-standing examples of computing education across computing disciplines include the Special Interest Group for Information Technology Education (SIGITE), the Journal of Information Systems Education (JISE), and the Information Systems Education Journal (ISEDJ). A few outlets of the many for non-computing disciplines are the Journal for Research in Mathematics Education, Journal of Statistics Education, The American Journal of Physics, and the Journal of Economic Education. Finally, there are numerous multidisciplinary resources for SoTL publications, such as the

American Educational Research Association (AERA), the International Society for the Exploration of Teaching and Learning (ISETL), and the International Society for the Scholarship of Teaching and Learning (ISSOTL).

Reflecting this desire across disciplines for research into teaching and learning with respect to educating students in professional ways of thinking and practice, particularly at the undergraduate level, a new area of academic inquiry was established to give “college and university teaching a place within a broader vision of scholarship that also included the discovery, integration, and application of knowledge.” (Huber, 2019) Entitled the Scholarship of Teaching and Learning, it grew out of the concerns expressed by Boyer (1990) that universities recognized only basic research as viable for promotion and tenure and would give the pursuit of excellent teaching considerably less importance. He advocated for the expansion of scholarship, in that “the work of the professoriate might be thought of as having four separate, yet overlapping, functions. These are: the scholarship of discovery; the scholarship of integration; the scholarship of application; and the scholarship of teaching.” (Boyer, 1990, p.16) It is this same perspective that enabled Meyer and Land’s seminal work (Meyer & Land, 2003).

SoTL is now an established field of research complete with international involvement, a professional society and annual conference of scholars ("International Society for the Scholarship of Teaching and Learning," 2004), and a number of peer-reviewed academic journals (Steiner, 2018).

A significant and rather overlooked group, as discussed in the next chapter, who should have input into essential concepts for students to master, are professionals in the field (Barradell, 2013). These professionals not only know what is required for the daily work, but also will be the colleagues of many of the future graduates and will, in some part, make decisions on whom – and from where – to hire, based on the preparedness of the students they encounter.

It is important therefore to all concerned with higher education that we identify those topics that, although perhaps initially difficult to grasp, students must surmount in order to progress in their ability to function professionally in the discipline.

In the education of students, within many academic areas if not all, there are concepts deemed critical for students to grasp and which, when incorporated, change their understanding in very significant and lasting ways, formative of how those in the discipline think and practice (Meyer & Land, 2003; Sorva, 2010). A relatively new theory of such concepts is that of Threshold Concepts (TCs), the background of which we explore in more depth in the literature review. The theory posits that such TCs exist and that they differ from other discipline concepts, even from some of what may be called core concepts, by the embodiment of certain characteristics. While not all of these characteristics are required to be present simultaneously, some do appear to be necessary in all cases. The following is a brief summary of five commonly listed characteristics (Alston, Walsh, & Westhead, 2015; Meyer & Land, 2003; Rhem, 2013; K. Sanders & McCartney, 2016):

Transformative – a changed view of the subject matter or even the discipline. This seems to be a required element of a TC, as deemed by most researchers. (Boustedt et al., 2007; Cousin, 2006; Davies, 2006; Eckerdal et al., 2006; Eckerdal et al., 2007; Land, Meyer, & Baillie, 2010; K. Sanders & McCartney, 2016; Sorva, 2010)

Irreversible – once learned, it is highly unlikely, if not impossible, to revert to the previous understanding; i.e., to unlearn. This appears to be an identified characteristic in most research work on TCs. (Boustedt et al., 2007; Davies, 2003, 2006; Eckerdal et al., 2006; Land et al., 2010; K. Sanders & McCartney, 2016)

Integrative – allows students to see connections or relationships to material within the discipline, which previously were undetected. Often this is seen as inherently connected to the transformative nature of a TC. (Boustedt et al., 2007; Cousin, 2006; Davies, 2003, 2006; Eckerdal et al., 2006; Land et al., 2010; K. Sanders & McCartney, 2016)

Troublesome – challenging to the student, but in a way that is more than simply difficult knowledge. Whether due to conceptually difficult material, a concept which is alien to the student, or simply counter-intuitive, a TC may bring such a novel view of previous understanding that it conflicts with their current interpretation, inhibiting transfer from one context to another, even if only momentarily. Often present but sometimes not recognized due to differing periods of resolution, some within a short timeframe.

Bounded – usually (not necessarily always) delineative of borders between conceptual spaces, and often disciplines. Clearly, there are limits to how much material a TC can integrate. However, a TC delineates principles understood by those within the discipline and not by those outside.

It therefore is an appropriate area ripe for study to investigate whether such concepts put forth as possible TCs are viewed as such by current professionals.

1.3 Contribution to the discipline

Education research in the computing disciplines has continued to expand and a number of research projects have addressed the discovery of TCs, principally in the area of Computer Science and often with regard to programming concepts (K. Sanders & McCartney, 2016). This research has been oriented towards student perceptions of subject material, which researchers subsequently examined for apparent conformity to TC characteristics. TC research in computing concerning instructors has been oriented toward inquiring of their opinions for candidate TCs. Very little research has occurred in the area of students in Web development, which encompasses teaching not only general programming concepts, but also an understanding of networks, databases, and security, in addition to the aesthetic issues of Web page design, all amid a variety of constantly changing technologies (Connolly, 2019). In the area of Web development, the work is significantly oriented towards instructors (Alston et al., 2015). As indicated earlier, an important instructional goal is to produce students who can think and practice as professionals. This study regards professionals as sharing some aspects with students (whether formally or informally) who have reached the goal of being a professional. They also share some aspects with instructors, as having experience with professional requirements. Yet almost no research has addressed Web development professionals on the subject (Dorn & Guzdial, 2010; K. Sanders & McCartney, 2016).

Virtually all previous research is qualitative, encompassing biographies, semi-structured interviews, verbal transcripts, and questionnaire methods (Alston et al., 2015; Moström et al., 2009; Walker, 2013). This paper explores a quantitative approach towards determining whether practitioners in the field perceive selected candidate concepts as TCs and how their educational and experiential backgrounds affect those perceptions. As indicated above, at least some of the characteristics listed are connected. Therefore, it was expected that, due to this interconnectedness, it could be difficult to develop clear measures distinct enough as to differentiate the presence of one characteristic versus that of another.

As a necessary outcome in seeking to quantify the association of constructs such as troublesome, transformative, irreversible, integrative, and bounded with the perception of a TC, this work has had to develop measures for each construct. Such development becomes a major contribution of this work as such a scale useful for quantitative work did not exist previously. By being able to identify measures for required attributes of a TC, this work will help to clarify some of the impreciseness for which previous researchers have criticized the theory (Barradell, 2013).

1.4 Research Agenda

The question of how to determine whether a given concept is in fact a TC is not fully established at this time, which is why TCs identified in the literature often are referenced as "candidate" TCs and why quantitative measures are needed. Zwaneveld, Perrenet, and Bloo (2016) say there is (as of now) no final answer as to the best source for TCs. They suggest that the best source "is to sit with students during the whole learning process" (p. 281), albeit a manifestly impractical approach. It is their contention that the second best option is to ask the students, which they reference as the primary source. This study regards working professionals as an extension of the student with an important additive: they have real experience to guide what knowledge truly is critical. This approach also addresses the concern of at least one set of researchers for allowing some elapsed time to check irreversibility, as well as their contention that progress will depend upon the addition of more involvement from professionals (K. Sanders & McCartney, 2016). Accordingly, this research seeks to extend work on a TC from a selected list of previously identified candidate TCs.

1.4.1 Problem Statement

In the quest to determine TCs in computing, the discipline of Web development is highly underrepresented,. Most prior research identified TC candidates having an orientation towards programming, which is applicable across multiple computing specialties. What research that does exist principally has been concerned with the opinion of instructors as to the concepts with which their students struggle in a manner indicative of possible TCs. The relatively recent study by Alston et al. (2015) is indicative. Student responses, however, may miss identifying some concepts as TCs, given that students do not have sufficient experience in the field to recognize the essential nature of a given concept or to know if an important topic was missing from their education. Experience, both in the discipline and in teaching, is an argument for using the determination of instructors as to what should constitute a given TC candidate. However, instructors may have difficulty identifying transformative material as they themselves have long since adapted to the candidate concept and could have difficulty identifying with the struggle students are likely to encounter. Although Zwaneveld et al. (2016) identify students as a prime source, using research results from students also may not be able to identify concepts as irreversible, given the proximity to the initial learning of the concepts. Finally, attempts to identify the perceptions of TCs by Web development professionals are essentially unexplored, although such professionals have the advantages of experience in the discipline, sufficient time from initial exposure to indicate whether a given concept is irreversible, and yet perhaps are not so distanced as to forget difficulties in assimilation of candidate TCs. This study identifies as a research issue the determination of whether a certain candidate TC is perceived as such by Web development professionals and the effects of educational and experiential background on said perception.

1.4.2 Research Questions

In order to address the identified research issue, we propose the following research questions:

1) To what degree is the candidate TC perceived as such by Web development professionals?
2) How are the perceptions of the TC influenced by the educational backgrounds of Web development professionals?
3) How are the perceptions of the TC influenced by the experiential backgrounds of Web development professionals?

1.4.3 Research Model

Beyond noting that the opinions of Web development professionals have had little representation in the determination of TCs, the choice of such professionals as research subjects is motivated by two factors relating to their time as professionals and their experience as students, respectively.

The first factor is that they have the advantage of experience in the field, unlike students but perhaps more like instructors. As such, they should have some relevant views on what topics are necessary to think and act as a professional; that is, what is needed to engage in professional discourse. In this respect, it is appropriate to ask whether their experience influences their perception of a topic regarded as a good TC candidate from earlier studies. Those studies that did look at instructors did not inquire as to the length of experience, either as an instructor or as a practitioner, or whether they had such work experience. The position of an instructor in the discipline presumes sufficient experience, both as an educator and as one familiar with critical concepts in the field. This study examined professionals who are working in the field and sought to determine if their experience, in computing overall and in Web development, specifically, influenced their perceptions. Given the emphasis upon providing real-world experiences for students, as delineated in both the 2008 and 2017 ACM Model Curriculum for Information Technology, experience must be considered as an important factor (Lunt et al., 2008; Sabin et al., 2017). Furthermore, as being experienced is an argument for the use of instructors in identifying TC candidates, it would appear likely that the experience of professional Web developers would impact their concurrence of those concepts as TCs. Determination of the influence of work experience on perceptions of TCs would be worthwhile, as this has not been addressed previously.

The second factor is their relative formal educational experience, or lack thereof. The only previously identified study utilizing working Web developers had few participants with a degree in a computing field (Dorn & Guzdial, 2010). Although a small study of only twelve participants, fully half of those with degrees were in areas not associated with computing. While there were some statements that academic work in computing could be advantageous, the general response was that such an education was not necessary, at least for Web designers working on the user-visible “front-stack” side, which most of these were (as opposed to “back-stack” developers working on server-based software.) Indeed, some had no degree. This raises the question of how formal education influences the perception of TCs. One might assume that such education would affect positively their perception of TCs, in that they should have both assimilated the concepts and be able to recall their experience of learning the concepts.

The Dorn and Guzdial (2010) study also had participants with a wide range of experience, ranging from 2 months to 15 years with given tools, and although not a study for examining TCs per se, there was no indication as to whether experience might affect perception of TCs. It may be that the passage of time and the gaining of expertise could ameliorate the troublesome and even the transformative impact of a TC on the learner. If this is the case, experience might correlate negatively with their perceptions of the “TC characteristics” of a given candidate TC.

This study sought to utilize a candidate TC, either identified in multiple previous studies as a candidate TC for computing more generally, which also would apply to Web development, or as a candidate TC specifically for Web development. This study further sought to determine whether Web development professionals perceived this candidate TC as a TC by posing a series of statements directed to various characteristics of TCs. This study then examined whether the experience or education of Web development professionals influenced their perceptions. The research model is presented in Figure 1.

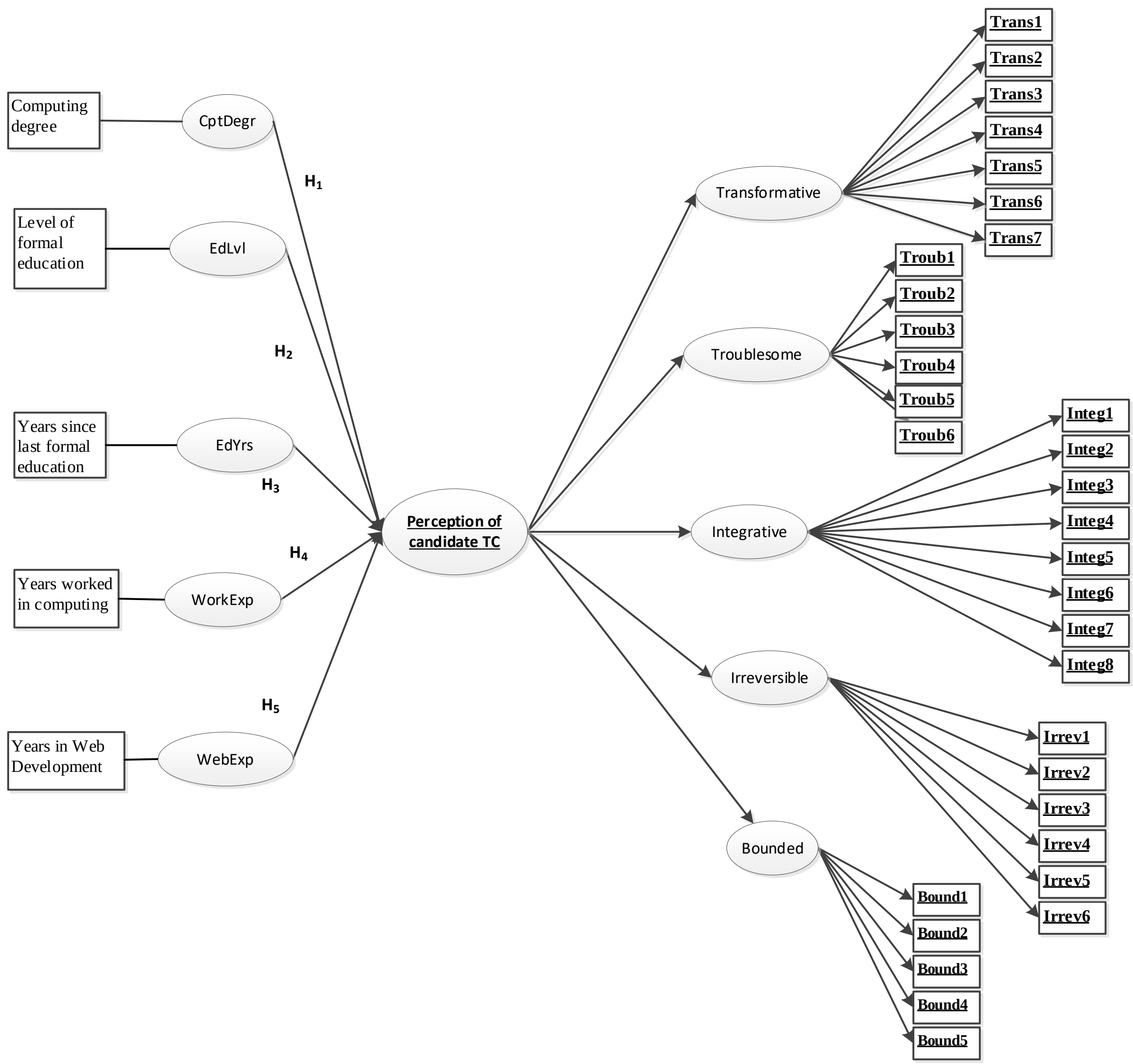

Figure 1. The Research Model

1.4.4 Hypotheses

Given the rationale explained in the discussion above, the intent of this study was to develop a new scale for the perception of a concept as a candidate TC by Web development professionals, with respect to five presented characteristics of TCs. This study takes the total number of years in computing, and the years in Web development, per se, as indicators of experiential background. As indicated earlier, experience may enable these professionals to have a greater appreciation for the requirements to "think and act like a professional".

As to educational background, this study also considers as indicators both the level of education received, in terms of degrees held, the type of degree (e.g., computing or otherwise), and the years since they last were in formal education (year of latest degree, or year last attended, if no degree.) Total years in Web development is particularly important when members of the population may not have had education experiences oriented around Web development, but only of other aspects of computing, which could include programming. It also is possible that they have had no formal education in computing but learned over time, informally and/or via professional development training programs.

The years of experience and/or since last in school may also contribute toward ameliorating the troublesome aspect of assimilating a candidate TC. Therefore, effects of experience and education may affect positively or negatively the perception of a candidate TC.

Given that no prior research has indicated the influence of education or experience of Web developers on their perception of TCs, and that either may have positive or negative results, competing hypotheses are presented. Rather than having hypotheses that indicate an influence but in an unknown direction and a null hypothesis of no influence, these allowed for testing the direction as either positive or negative. If, for example, a hypothesis simply stated that an influence exists (H_a: $\mu_a \neq 0$) and the null hypothesis being that there is no influence (H_0: $\mu_0 = 0$), rejecting the null hypothesis only says there is an influence, but nothing about its direction. Having competing hypotheses for a positive effect (H_0: $\mu_0 \leq 0$, H_a: > 0) and a negative effect (H_0: $\mu_0 \geq 0$, H_a: < 0), allows for showing the direction (rejecting the null) whether positive or negative. Therefore, the following hypotheses were proposed:

H_{1a}: For Web development professionals, having a degree in computing has a positive correlation to their perception of the candidate concept as a TC

H_{1b}: For Web development professionals, having a degree in computing has a negative correlation to their perception of the candidate concept as a TC

H_{2a}: The level of formal education of Web development professionals has a positive correlation to their perception of the candidate concept as a TC

H_{2b}: The level of formal education of Web development professionals has a negative correlation to their perception of the candidate concept as a TC

H_{3a}: For Web development professionals, the years since they last were in formal education has a positive correlation to their perception of the candidate concept as a TC

H_{3b}: For Web development professionals, the years since they last were in formal education has a negative correlation to their perception of the candidate concept as a TC

H_{4a}: For Web development professionals, the years they have in computing has a positive correlation to their perception of the candidate concept as a TC

H_{4b}: For Web development professionals, the years they have in computing has a negative correlation to their perception of the candidate concept as a TC

H_{5a}: For Web development professionals, the years they have in Web development has a positive correlation to their perception of the candidate concept as a TC

H_{5b}: For Web development professionals, the years they have in Web development has a negative correlation to their perception of the candidate concept as a TC

1.5 Summary

A brief discussion of Web development as a discipline was presented and some challenges with respect to developing applications for a wide audience via the Web were noted. A relatively new theory in education research, TCs, was summarized, holding promise to highlight those particular concepts necessary for students to progress in the discipline towards thinking and practicing as a professional. It was noted that previous research was qualitative, seeking to establish candidate concepts believed to meet the requirements for a TC. Five characteristics often observed for candidate concepts were highlighted. Subsequent discussion of approaches used in research for computing education indicated a lack of research in the area of Web development, particularly with respect to professionals in the field. It was suggested that new scale development was needed to be able to determine quantitatively whether a given candidate concept is viewed as meeting minimal criteria as a TC within the target group.

Finally, this study presented the hypotheses for proposed new research oriented towards verifying the perceptions of a selected candidate TCs by a neglected stakeholder group, professional Web Developers, and how those perceptions might vary by educational and experiential background.

CHAPTER 2: LITERATURE REVIEW

This chapter explores the Threshold Concepts Framework, which acts as the theoretical basis for this study. This framework is relatively recent, as underpinning research began in 2001 in the United Kingdom, therefore new research and theory development are continuing. Central to this study are the foundational dimensions of TCs and their appearance in Web development. This study explores their origin; characteristics and how these relate to each other; critiques; identification in academic fields, particularly computer science; research on TCs in Web development, and how this work intends to extend that research.

2.1 Origins of Threshold Concepts

The idea of TCs has roots in conceptual change (Carey, 1991), of which there exists a considerable literature, especially in the fields of cognitive and educational psychology. The theory also has connections with social learning theory, particularly with communities of practice as discussed in the work of Wegner (1998). From the beginning, it has had a close relationship to the Variation Theory of Marton, Runesson, and Tsui (2004), although a somewhat different model was proposed by Meyer, Land, and Davies (2008).

The TCs framework originated from a project in the United Kingdom, funded by the Economic and Social Research Council as a part of its Teaching and Learning Research Programme (TLRP). The research project, Enhancing Teaching/Learning Environments in Undergraduate Courses ("ETL," 2005), was designed to support 28 course teams across 15 departments teaching undergraduates, and aimed at improving the quality of learning in higher education ("TLRP," 2009). The project team involved individuals from three universities in the UK (Edinburgh, Durham and Coventry) with various collaborating institutions clustered around each of the universities. The project sought to develop subject-specific conceptual frameworks for guiding the development of teaching-learning environments. They used these conceptual frameworks to integrate research findings from the knowledge of academic staff with institutional and national standards for high quality teaching and learning. The intention was to evaluate the applicability of certain key concepts in high quality outcomes of learning. The key concepts as defined were:

- *Teaching-learning environment,* including the full range of teaching, assessment, learning support, facilities, resources, and administration, concentrating on those expected to influence most directly the quality of learning.
- *Constructive alignment*, meaning the 'fit' of the course objectives and the teaching, learning, and assessment procedures utilized, while focusing on the development of conceptual understanding.
- *Ways of thinking and practicing in the subject (WTP)*, indicating not just approaches to studying but the specific skills and thinking processes students of a given subject area need to develop on their way to becoming professionals in their field.
- *Troublesome knowledge and threshold concepts*, centering on topics and on ways of thinking that become a threshold to additional learning.

In Phase 2 of the program, Meyer and Land (2003) introduced their seminal paper on TCs. They initiated the idea of a *threshold concept* to differentiate between core concepts that embody "seeing things in a new way" and those that do not. The implication here is that all TCs are core concepts, but not all core concepts are TCs.

For this UK project, there were four areas chosen as subjects to cover academic disciplines and professional areas (electrical engineering, cell and molecular biology, business economics, and history.) Initially this was to be five, but there were no results posted for the topic of media and communications ("ESRC," 2005).

2.2 Characteristics of Threshold Concepts

TCs are discipline-specific core concepts that have a higher requirement for the learner to internalize. After discussions with practitioners across a range of disciplinary areas, as noted earlier, Meyer and Land (2003) identified five characteristics that they believed defined TCs: 1) transformative, 2) probably irreversible, 3) integrative, 4) often bounded, and 5) potentially troublesome. Although discussed in the conclusion of their work, liminality was not included explicitly in the list at that time. More recently, however, Meyer and Land (2005) and Land et al. (2010) expanded the list to consider discursive, reconstitutive, and liminal. Each of these is discussed in more detail below. Such ongoing qualitative research seeks to determine possible additional characteristics. However, as subsequent characteristics appear to be inherently a part of (or at least grounded in) the original, none have had any subsequent quantitative studies, and to keep to a manageable size, this study limited inquiry to the original five characteristics, as they have continued to be referenced since the seminal paper of Meyer and Land (2003).

2.2.1 Transformative

Meyer and Land (2003) stated that TCs are something distinct from what we normally might call core concepts, in that while a core concept may require understanding to help the student in comprehending the subject, “it does not necessarily lead to a qualitatively different view of subject matter.” (p. 4) The distinction is between core concepts that epitomize the necessity of seeing things in a “new way” versus those which have no such requirement. Without this transformation, the learner cannot progress in the subject area. The transformation from the prior conceptual understanding to the new conceptual space causes a significant shift in the perception of a subject, whether in whole or in part. The transformation need not be immediate but may happen over a substantial timeframe. Moström gives an example of a student making a lasting behavioral shift in an instant:

> *Subject 7*: My Teammate suddenly asked me 'what are you doing?' I was just copying and pasting some code . . .He simply stated, 'Whenever you want to copy [and] paste, you should be using a function.' (Moström et al., 2009, p. 184)

Moström then provided an example where the transformative shift to a different view took place over a longer period of time:

> *Subject 21*: My code became very long and hard to understand . . .After realizing that this was not the right way to go about making the program I thought back to one of the first lectures I was given on Object Orientation. It then dawned on me that the classes were representing real world objects, the dog class could be thought of as an actual dog . . . (Moström et al., 2009, p. 184)

Regardless of the time involved, without a changed view, it would appear that a concept may be considered to be a core concept, but not to be a TC. The transformative aspect of a TC thus is considered a required characteristic (Flanagan, 2015). Wilkinson (2014) posits that the definition of a TC "may" contain any or all of the characteristics listed here, and therefore it is difficult, if not impossible, to state that any one characteristic is required in all cases. In fact, the theory has undergone some changes in enumerated characteristics from that in its initial formulation, and these may continue to evolve with additional research. However, transformative does seem to be present where TCs are discussed, and this does appear to be accepted by almost all as an "essential property"(Baillie, Bowden, & Meyer, 2013).

2.2.2 Irreversible

Once learners reach comprehension of a TC, a point where transformative learning has occurred, it is unlikely, or only possible with significant effort, that they could undo their new understanding and return to their previous belief of the concept, and even then, not return completely to their original state. The new understanding possibly may appear so simple that they wonder why they had such difficulty in the first place. "Transformative learning involves experiencing a deep, structural shift in the basic premises of thought, feelings, and actions. It is a shift of consciousness that dramatically and irreversibly alters our way of being in the world." (Morrell & O'Connor, 2002, p. xvii) This characteristic may explain the difficulty an expert in a field may have when trying to understand the struggles of learners to grasp a concept. The expert has long since crossed the threshold that the learner is encountering perhaps for the first time.

2.2.3 Integrative

Acquisition of a TC not only transforms thinking but also enables the learner to integrate basic concepts through a grasp of vital interrelationships previously undiscerned; that is, exposing an interrelatedness previously hidden. The learner begins to *integrate* disparate aspects of their learning, to "think" like a practitioner in the field of their study.

> "The integrative aspect of a threshold concept presents distinctive problems for learners who are studying a subject (such as economics) as part of their degree. Students who do not think of themselves as 'learners of economics' are likely to face particular difficulties in grasping concepts that bind together aspects of a subject that may seem quite disparate to a novice. This problem arises because the acquisition of such concepts (e.g., opportunity cost, price and value, equilibrium) is intrinsic to grasping the ways in which economists 'think' and practice." (Land, Cousin, Meyer, & Davies, 2005, p. 54)

Without this integrative aspect, whether majoring in the field or not, conceptual difficulties not only will continue but also may well increase. Therefore, Meyer and Land (2003) listed *integrative* as a property without the “probably”, “often”, or “potentially” qualifiers. Subsequently, Meyer referenced E.-J. Park and Light (2010), where they characterized integrative as “indispensable” (Meyer, 2016, p. 464).

Davies (2003) stated that as teaching academics, one not only looks to discover TCs but “Given the importance of ‘integration’ to the threshold concept it is also important to examine the degree to which a threshold concept is evident in students’ thinking about different phenomena.” (p. 13) That is to say, instructors should note whether the learners are integrating concepts in order to think and practice like a professional in their field.

2.2.4 Bounded

A TC can integrate only so much. Possibly often, but not always, a TC has terminal frontiers which border thresholds into new conceptual spaces. These boundaries at times may delineate borders between academic disciplines.

> “Within the field of Cultural Studies a threshold concept that has to be understood early is the breakdown of the barrier between high and popular culture. This is fundamental to the Cultural Studies approach. This is a significant departure from practice in English Literature where that concept not only doesn’t really exist but if it did (i.e. if you crossed that threshold) it would undermine the discipline of Eng. Lit. itself.” – personal communication from Siân Bayne, 2002, quoted in Meyer and Land (2006b, p. 8).

It is not surprising that academic disciplines with overlapping subject material would share concepts that qualify as TCs. That Chemistry and Physics relate closely so to have placement in the same department in some schools is unsurprising. Yet Chemistry and Biology typically are in distinct departments, although these disciplines do overlap in key areas. Biological studies require an understanding of certain chemical processes and the resulting overlap may give rise to single courses or entire major areas, such as Biochemistry. Distinct disciplines do not have to have such obvious connections in order to have some TCs in common. For example, informatics applies to healthcare (health informatics, nursing informatics), life sciences (bioinformatics), and the arts (2-D and 3-D animation and visualizations; sound recording, mixing), in addition to business environments. However, one might well expect to have many, if not most, discipline TCs bounded by the particular academic field.

2.2.5 Troublesome

TCs may comprise, lead to, and intrinsically represent conceptually difficult knowledge. Perkins referred to this as *troublesome knowledge*; that is, counter-intuitive, alien to one's culture or discourse (Perkins, 1999). Meyer and Land (2006a) suggested that knowledge which transforms "probably should be troublesome … but that does not mean it should be stressful or should provoke the kinds of anxiety, self-doubt and frustration that can lead students to give up." (p. xiv)

Troublesome knowledge, as a component of a TC, serves an *instigative* function, unsettling prior cognition, and pushing the learner into a state of liminality (from the Latin *limen*, meaning "threshold"; a suspended state of understanding which approximates to a type of mimicry or lacking of authenticity.) Invalidating a "common sense", or intuitive, understanding can inhibit mastery of a TC, as such a collapse can be emotionally uncomfortable. The learner may not be willing to let go of their current way of understanding even if it is limited. The process of letting go and integrating new knowledge involves an ontological shift because "We are what we know." (Land, 2015, p. 20)

New insights may be exciting but may also involve a sense of loss. According to Land, “Perhaps the threshold concept is so troublesome not because the concept is so difficult but because it challenges the learner’s understanding of its component concepts and this is why it acts as a check point for the learner’s progress.” (Land, 2015, p. 25) Perkins suggests that concepts in a discipline functionally categorize knowledge, bringing both considerable gain as well as particular problems:

> “As [a student] learns a fresh conceptual system, a new world emerges. The pre-Freudian self does not look the same as the Freudian self with its id, ego, and superego … [but the] categorical function of concepts also brings its distinctive troubles. [The student] is likely to find a novel parsing of the familiar world confusing and confounding, for instance mixing up id and ego, or mass and weight.” (Perkins, 2006, p. 41)

Perkins also suggested that some TCs give rise to being troublesome not from the concepts themselves, but from how they fit together to create an underlying “conceptual game”, or “epistemic game”, for transforming learner understanding at a deep level. This requires the learner to comprehend the (often tacit) “games of enquiry”, or “ways of thinking and practicing” (WTP) within the domain of specific disciplines. As disciplines are more than a bunch of concepts, each has a characteristic *episteme*, “a system of ideas or way of understanding that allows us to establish knowledge.” (Perkins, 2006, p. 42)

It always is important to keep in mind that as each discipline may have troublesome topics for new learners, troublesomeness alone does not indicate that the topic is a TC. However, it may indicate an area of inquiry as to the existence of a TC.

2.2.6 Reconstitutive

Once in a liminal state, the learner must integrate newly acquired knowledge into their conceptual schema. This requires a reconfiguration of their conceptual understanding. They simultaneously must discard their earlier primitive understanding. Taken together, this “integration/reconfiguration and accompanying ontological/epistemic shift can be seen as *reconstitutive* features of the threshold concept.” (Land et al., 2010, p. xi) This reconstitutive feature is an implication from the transformative and discursive aspects of TCs. It is more likely that others apart from the learner will recognize the change, and that the change takes place over a period of time rather than in one moment.

2.2.7 Discursive

By the term discursive, Meyer and Land indicate that crossing a threshold results in an enhanced and extended use of language.

> "It is hard to imagine any shift in perspective that is not simultaneously accompanied by (or occasioned through) an extension of the student’s use of language. Through this elaboration of discourse new thinking is brought into being, expressed, reflected upon and communicated. This extension of language might be acquired, for example, from that in use within a specific discipline, language community or community of practice, or it might, of course, be self-generated. It might involve natural language, formal language or symbolic language." (Meyer & Land, 2005, p. 374)

Barradell (2013) argued that since TCs are bounded and are associated with the WTP of a discipline, that the discourse on what concepts are TCs should go beyond academics and learners in the classroom to include professionals in the field. This inclusion would identify what the discipline and its workforce require rather than only instructor/student views. Without this inclusion, identification of TCs is missing a very relevant perspective.

As indicated earlier, one's discourse undergoes a change with new words (symbols, signs) introduced, or old words re-assigned, to describe new concepts. Land, Rattray, and Vivian (2014, p. 204) used the metaphor of a "cognitive tunnel" which they illustrated diagrammatically, for introducing the learner to a new concept through introduction of a new signifier, or sign, including terminology and symbols (see Figure 2), and for the introduction of a new meaning to an existing sign (see Figure 3.)

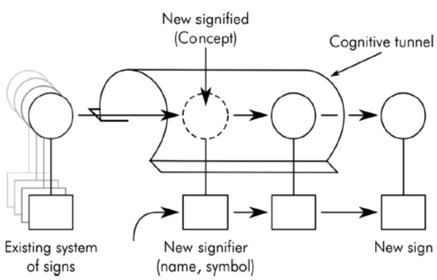

Figure 2. Introducing a new concept
(Figure used by permission)

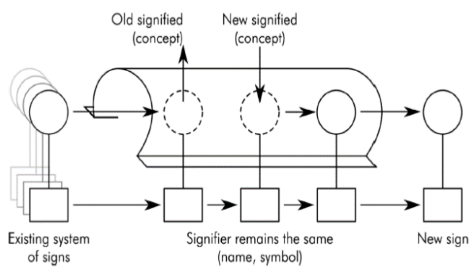

Figure 3. Introducing a new meaning
(Figure used by permission)

They admitted, however, that it is problematic to attempt capturing ideas such as TCs and liminality, as well as troublesome knowledge, in a diagrammatic fashion.

2.2.8 Liminal

Learners having difficulty comprehending a TC may find themselves in a state of liminality, a "stuck place" where there is some level of incomplete understanding yet they are unable to proceed for some period of time (and perhaps never) to full integration of the TC. In this state, learners may exhibit mimicry rather than truly comprehend. That "aha" moment when the details come together in understanding a concept may appear to be due to a sudden insight, but often a long time in the liminal space precedes it. If a truly sudden insight happens, without much time in the liminal space, usually this is due to a deep understanding in a related area (Eckerdal et al., 2007). McCartney et al. (2009) indicate that the liminal space can include partial understanding of the concept, even though the learner has not yet mastered it. In discussing the liminal state, Meyer and Land drew from seminal works in ethnographical studies into social rituals by van Gennep (1960) and Turner (1969). Examples could be rites of passage initiating boys into manhood, or that liminal space/time between the “threshold” ceremony of a wedding and the “transformative” realization of being married, referenced in our own culture as a honeymoon. Turner described the transitional time and space within which their subjects conducted such rites with the term “liminality”, which Meyer and Land subsequently to characterize the difficult time for students between exposure to and realization of a TC (Meyer & Land, 2005). In both examples presented here, there is a transformative aspect, the subject(s) acquire(s) new knowledge and a new identity or status within a community, and the transition frequently is troublesome, problematic, and often humbling. Finally, once entered, there is no return to the pre-liminal state. While some learners never find their way completely through the liminal space, having the new concept introduced means they are forever changed. “It would appear however, that once the state of liminality is entered, though there may be temporary regression, there can be no ultimate full return to the pre-liminal state.” (Meyer & Land, 2005, p. 376)

As learners progress through the liminal state, however quickly or haltingly their path may be, they converge on the way of understanding (or WTP) appropriate for their discipline. Learners do take differing times and routes, and the sticking points are not always the same in place or in number. “The most important practical observation from this work may be that different students take different routes through the liminal space, with the possibility of getting stuck at multiple places." (Eckerdal et al., 2007, p. 9)

Learners find their way through this liminal state by a multiplicity of strategies: reading, searching the Internet, learning from others (a peer or another faculty member), obtaining more examples from their instructor, visualizations, utilizing lots of practice, decomposing into smaller parts, or even looking for the "larger picture" (McCartney, Eckerdal, Moström, Sanders, & Zander, 2007).

Various metaphors are in use to describe TCs and the entrance into a liminal state. Meyer and Land (2005) use a "doorway' or "portal" metaphor, stating, "A threshold concept can be considered as akin to a portal, opening up a new and previously inaccessible way of thinking about something." Regardless of the particular metaphor, this concept of "entering in" is particularly true as regards the transformative aspect leading the learner into a liminal state and, hopefully, into an integrative cognition of the concept at hand. Land et al. (2010) provided a relational diagram of the features of TCs (see Figure 4), wherein they showed pre-liminal, liminal, and post-liminal modes of a TC.

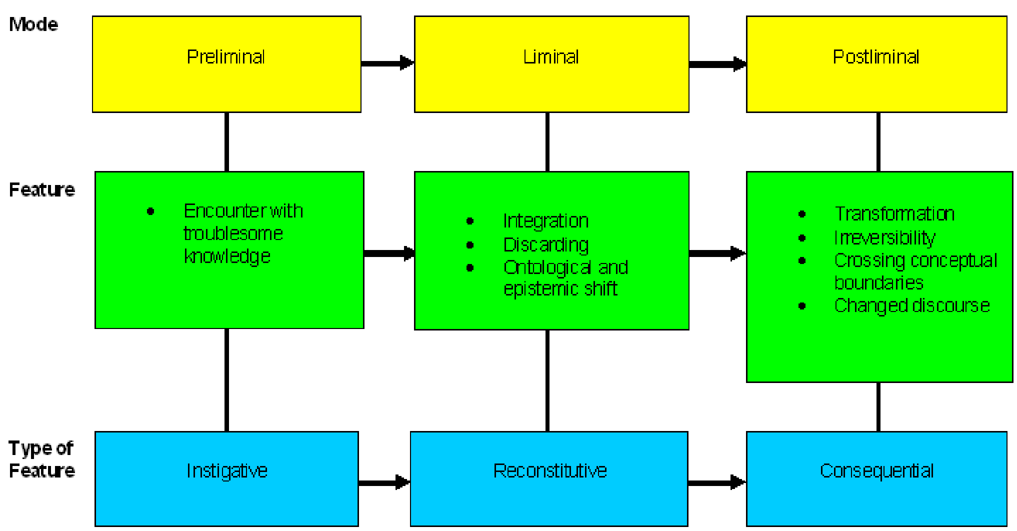

Figure 4. A relational view of the features of threshold concepts.
(Land, Meyer, Baillie, 2010, p. xii)

This diagram has the strength of illustrating the learner's encounter with a TC and the features relative to the pre- liminal, liminal, and post-liminal modes.

Encountering troublesome knowledge is a feature of a pre-liminal mode (or state.) This study indicated in an earlier section that such an encounter is instigative, and their diagram illustrates this as well. When the learner encounters troublesome knowledge, they move into the liminal state, where they begin the process of integrating this new knowledge. During the liminal state, this diagram shows the nature of the reconstitutive aspect of TCs, with the integration of new knowledge, discarding of previous incorrect or limited beliefs, and the causing of an ontological and epistemic shift, which leads to the transformation characterizing the post-liminal mode. Lastly, the diagram shows the post-liminal mode as being when the learner has transformed, crossed the conceptual boundary, and changed their discourse.

Yet Land et al. (2010) indicated that this view "has an overly rigid sequential nature." They suggested that an illustration of oscillation and recursiveness needed overlaid on their diagram. Flanagan (2015) illustrated this (see Figure 5) through his link on liminality where the previous diagram appears and then this second one overlays it as the user moves their mouse pointer over the first diagram.

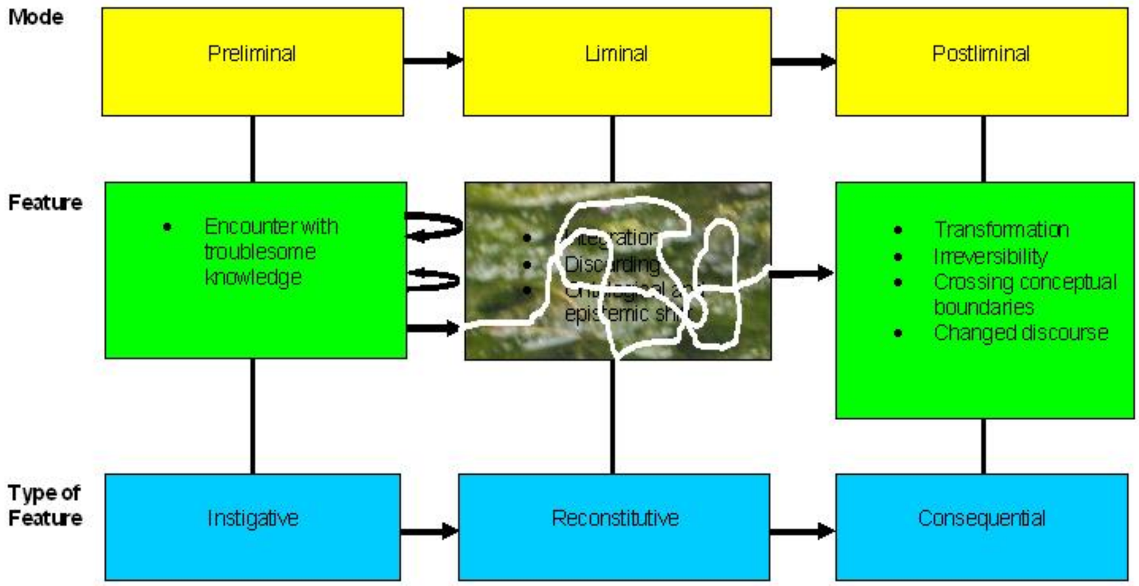

Figure 5. Accessed via – http://www.ee.ucl.ac.uk/~mflanaga/popupLiminality.html

The meandering path through the center feature set, along with the returning path back to encounters with troublesome knowledge, indicate that the path from a pre-liminal state through liminal to a post-liminal state is neither straight nor a foregone conclusion. Some may wander in a liminal state for an extended period, fail even to enter, or return to a pre-liminal, yet not entirely original, state.

Cousin commented, “In short, there is no simple passage in learning from ‘easy’ to ‘difficult’; mastery of a threshold concept often involves messy journeys back, forth and across conceptual terrain.” (Cousin, 2006, p. 5)

2.3 Discussion on Relationships of Characteristics

Davies and Mangan (2005) described the interconnectedness of the first five characteristics of TCs. They state that bounded and troublesome are derived from the first three, transformative, irreversible, and integrative, and that these first three are interwoven. A TC that integrates prior understanding changes the perception of the student’s current understanding, and thus is transformative. Being transformative affects more than just cognition. When the learner grasps the concept and has that “aha” moment, it is both an emotional and a cognitive experience. As the TC incorporates the student’s understanding concerning different phenomena, it is more likely to be irreversible once acquired, for it would be greatly disturbing to the student’s way of thinking if they were to abandon the TC.

As to the last two characteristics which they investigated, bounded and troublesome, the more a TC transforms a student’s way of thinking, the more likely it is to be troublesome, as it requires a reconfiguration of previous understanding. A TC helps define subject boundaries, as it integrates a given set of concepts, theories and beliefs. The stronger the integration, the stronger the boundaries will appear, and conversely, the looser the integration, the more ill-defined the boundaries, opening them to debate.

Although not addressed by Davies and Mangan (2005), the characteristics of discursive, reconstitutive, and liminal also are related. Meyer and Land (2005) underscore how difficult it would be to have new thinking from the acquisition of a TC, without a simultaneous extension in the use of language, an elaboration of discourse. They note that with such changes come a repositioning, or reconstitutive, effect on the learner’s identity, emphasizing the interrelatedness of identity, thinking, and language. One might say, “You are what you think and speak.” Finally, one can hardly expect such changes in thought, language, and even identity to come without encountering some concomitant sense of crossing a threshold or boundary into unknown territory.

2.4 Threshold Concepts and Traditions of Conceptual Origins

Carey (1991) discussed the ongoing debate for origins of human concepts between the “conceptual change” and the “enrichment” views of cognitive development, defining enrichment as consisting “in forming new beliefs stated over concepts already available.” (p. 460) The fundamental premise is that previous understanding makes incorporation of new concepts easier. This often results in the visualization of new concepts as “building blocks”, where the curriculum emphasizes “layers of understanding”. This implies that one concept is more complex than is another because of the necessary additional layers of understanding (Davies & Mangan, 2007). After giving two examples of enrichment views demonstrating the argument of knowledge acquisition depending on innate concepts or innate primitives, Carey (1991) stated, “the conceptual change position stakes out a third possibility, that new concepts may arise that are not definable in terms of concepts already held.” (p. 460) Carey further defines concepts as “structured mental representations”, which was in agreement with theorizing in cognitive psychology.

Carey proposed three types of conceptual change, wherein reworking prior knowledge surfaces a new concept:

1) differentiation of a prior concept into two or more concepts;

2) merging of two or more prior concepts into a single, new concept; and

3) reanalysis of properties as relations, moving from a definition of a concept in terms of properties to one in terms of relationships.

The previous description of TCs places them within the conceptual change tradition rather than that of the enrichment tradition.

2.5 Critiques of Threshold Concepts

Critiques of TCs certainly are appropriate to validate the theory. Those who support the theory with a valid perspective do not demand blind acceptance. However, the effort to identify TCs within various disciplines is ongoing for as Perkins (2008) states, “We need at least a provisional commitment to one or another framework, leavened by an appreciation that it is subject to challenge from other viewpoints and to ongoing reconstruction." (p. 6)

The critiques of TCs have not been many. In the relatively short time since the formulation of the TC theory, it has spread across a wide variety of disciplines. That is in itself a critique, as some believe that sufficient critical scrutiny of TCs is lacking. Wilkinson (2014) expressed this very concern, among others. However, Wilkinson's concerns were contested by Townson, Lu, Hofer, and Brunetti (2015) in their defense of the TC framework. However, Barradell (2013) argued that the ready and broad acceptance of TCs

> "has meant that aspects of the discussion around threshold concepts have not necessarily been undertaken with the rigour they perhaps should, and that a number of important questions remain unanswered. For example, how many of the five characteristics should a concept possess to be regarded as a threshold concept? Are some characteristics more important than others? If a concept is troublesome and integrative but not transformative, is it still a threshold concept?" (p. 266)

Barradell raised some good and important questions. However, in reading of the seminal work of Meyer and Land, it appears that Barradell may have missed, in the language of the original formulation of TCs, the requirement for the transformative characteristic as a minimum. Baillie et al. (2013) concurred with Meyer and Land's original list, that the transformative aspect of a TC is an essential property, a statement to which this study subscribes. Rowbottom (2007) at first appeared to make a similar oversight ("But they only say that '[a] threshold concept . . . is likely to be' each") even though he later allowed an assumption that this as an "essential property", yet then argued that all concepts potentially are transformative (p. 263). The issue for TCs has to do with concepts in a particular field of endeavor (although they may apply to other fields, particularly if the fields are related) and if the concepts are of such import that not grasping them would inhibit further success. He did raise some legitimate issues, such as Meyer and Land not defining the term “concept” in their original work. He later contended, “… being transformative is arguably an extrinsic property, rather than an intrinsic one. Specifically, what is transformative for Mr. A need not be so for Mrs. B, because this depends on the conceptual scheme (or system of concepts) initially possessed by each.” (Rowbottom, 2007, p. 267). It would appear that prior experiences of one person over that of another may make a given TC more easily understood, but that does not rule out an intrinsic transformative nature of a TC. It only indicates that the first person was “further down the road” to understanding, having had experiences helping to form their more accommodating conceptual scheme. This might contribute to why some have little time in a liminal state for a given TC while others take much more time.

Another concern may be the granularity of a given TC. Sorva argues that some candidate TCs (e.g., state) are really more general “fundamental ideas”, “teachable at all levels” but not transformative, while others he labels as more specific “transliminal concepts”, lying beyond the actual TC but which “`lures' the student to and across a threshold concept” (Sorva, 2010). Rountree and Rountree (2009) indicate that not only is it difficult to come to agreement on granularity but that there may be hierarchies of TCs at varying levels of granularity. Clearly, the appropriate granularity of a given TC is an issue that has yet to be resolved. Therefore, for the purpose of this study, we consider candidate TCs which have been around for some time, in one form or another, and/or which have been specifically applied to Web development at some time.

Noting shortcomings with common methods used to identify TCs, Shinners-Kennedy and Fincher (2013) decided that the methods they used brought them to a “dead end”. Concluding that *hindsight bias* and attempted recall of emotionally laden events (e.g., encountering a TC), unexpectedly discovered in their own research of TCs, gave them pause to the consideration of retrospective accounts of learning a TC, which led them to consider other approaches; specifically, to the use of a two-dimensional grid for content representation (CoRe) to capture and represent “pedagogical content knowledge” (PCK). Rather than a repudiation of TCs per se, they saw CoRe as a methodology for TC discovery. Zwaneveld et al. (2016) addressed these concerns of hindsight bias and emotional load, disagreeing on theoretical and practical grounds.

Some have suggested that the scarcity of critical analysis makes the development of the actual concepts within a particular field languish. Morgan discussed from the field of Library Science as a case in point that “groupthink” and popularity greatly influenced the utilization of the TCs approach in the reworking of information literacy standards.

> “Was a ‘threshold concepts approach to the revision’ chosen because it was best, or because it was most popular at the time? Its popularity is hard to dismiss as a coincidental matter. Popularity is not itself a negative attribute, and I don’t mean to suggest that a popular paradigm cannot also be a valuable one. In this particular case, however, the popularity of the “approach” and the paucity of criticism of it have resulted in a curious stagnation in the development of information literacy threshold concepts *themselves*.” (Morgan, 2015, p. 282)

While this view is understandable and possibly true, it does not make the TC framework incorrect or less valuable. Furthermore, identifying TCs in a discipline is not necessarily an easy or direct issue (Davies, 2006; Davies & Mangan, 2007).

2.6 Identification and Usage of Threshold Concepts in Various Academic Fields

In spite of the relatively short history of TCs since the initial project began in the UK, a diverse collection of academic fields has utilized the TCs framework. Specific TCs may differ from one field to another. However, as the focus of the original research was on characteristics of strong teaching and learning environments across diverse disciplines, it is not surprising that the framework appears to be applicable to all. All academic fields have concepts necessary for a professional level of performance, but which are difficult and unsettling to learners. The number of disciplines with papers regarding the use of TCs is too many even simply to list here. This section briefly examines the TCs as applied to a selection of fields, including Computer Science, and where applicable, how they might relate to TCs in Web development, about which very little research exists today.

2.6.1 Economics

Economics was the first field to have the TC framework applied, as a part of the ETL project ("ETL," 2005). Economics often is an example subject in discussions regarding TCs, perhaps due to researchers depending on the seminal work or to the widespread inclusion of Economics in general education, accompanied by a typical sense of confusion on the part of non-economists. For Economics, some typical examples of TCs are *opportunity cost* and *elasticity*; opportunity cost being the value of the most valued in the group of rejected alternatives; elasticity being the degree consumers change their demand (or producers, their supply) due to a change in price or income.

Davies & Mangan proposed three types of conceptual change in Economics: basic, discipline TCs, and modeling concepts (Davies & Mangan, 2007, 2011). *Basic* concepts provide for categorization of phenomena in ways necessary for the deployment of TCs. Although students may become stuck with a basic concept, Davies and Mangan concluded such a concept does not employ significant integration and does not rise to the level of a TC. *Discipline thresholds* are attendant with ways of practicing the subject; that is the procedures specific to the discipline (Economics) and used in the construction and analysis of arguments - the use of models in the case of Economics. As an example, Davies and Mangan mentioned the use of "comparative statics", involving the use of comparative positions of equilibrium and the concept of "ceteris paribus" (all other factors remaining unaltered.) The third type they proposed, *modeling concepts*, are the learning of how to select, change, and test economic models, which is central to undergraduate economics education.

2.6.2 Education

As the TC theory arose from researchers in Education (Meyer & Land, 2003), it is not surprising that in each academic field TCs are considered central to learning (and therefore educating) in that discipline. As one might expect, there are a plethora of papers on TCs written by those from this discipline. Many of the works from educators are done as research in higher education investigating TCs in other disciplines. Some works are more general than discipline-specific. In their research on TCs in graduate work, specifically graduate research, including at the doctoral level, Kiley & Wisker concluded: "there are possibly six threshold concepts in research education, that is, six major conceptual challenges for those learning to be researchers, these are the concepts of:

- *Argument*
- *Theorising*
- *Framework*
- *Knowledge creation*

- *Analysis and interpretation*
- *Research paradigm*

Perhaps just as important was their indication that stakeholders (students as well as supervisors) "can identify learners' achievements … in terms of working at the level appropriate for graduate research" (Kiley & Wisker, 2009, p. 439).

Others works were more discipline-specific, such as: *research question* and *research objectives* in teaching research methods in business (Maciocha, 2014).

2.6.3 Additional Academic Fields

Some typical examples of TCs in other fields that could be transformative to learners might include:

Physics: *force, entropy*, *uncertainty in measurement,* or *gravity* (versus the center of gravity) (Meyer & Land, 2003, 2005; Wilson et al., 2010)

Accounting: *depreciation* (Meyer & Land, 2005, p. 374)

Business Leadership: *situational leadership* or *shared leadership* – as opposed to common concepts of heroic leadership (Yip & Raelin, 2012)

Engineering: *Laplace Transform, Equilibrium of Moments*, *properties of fluids* (Male & Baillie, 2011)

Biology: *thinking at submicroscopic and subcellular levels* for understanding of cellular metabolic processes (Ross et al., 2010)

Law: *precedent* or *legal reasoning* ("thinking like a lawyer") (Weresh, 2014)

Mathematics: *limit* or *complex numbers* (Meyer & Land, 2003).

It becomes intuitive that each field has its own set of discipline-based TCs, even though there may be overlap at some level.

2.6.4 Computer Science

Boustedt et al. (2007) using qualitative research techniques with instructors and students investigated whether there are TCs in CS and concluded in the affirmative. They identified *object-oriented programming* (OOP) and *pointers* as TCs in Computer Science, concluding that these met the criteria. As support to the notion of OOP meeting TC criteria, they referenced Luker's assertion that OOP is "a true paradigm shift, which requires nothing less than a complete change of world view." (Luker, 1994, p. 56) They did allow that these TCs might be broad areas within which may exist TCs that are more specific and suggested more specific TCs might include *concurrency* and *decomposition* (Boustedt et al., 2007, p. 508). Perhaps more specifically, K. Sanders et al. (2012) after conducting semi-structured interviews with students indicated that, particularly in CS, there might be Threshold Skills in addition to TCs. They suggest that knowing "what" are the TCs whereas knowing "how", the actual ability to do, are Threshold Skills.

Eckerdal et al. (2006) in a review of related work in computing concluded that *abstraction* and *object-orientation* are possible TCs, indicating that both meet the criteria. They referred to "the ability to abstract" as a key skill in Computer Science and then referenced findings from Or-Bach and Lavy (2004) whose work presented results from a student exercise to identify that abstraction is key to understanding OOP and that it is "a higher-order cognitive skill difficult for students to conceptualize." (Eckerdal et al., 2006, p. 106) With respect to object-orientation, they provided evidence that this concept is irreversible, troublesome, transformative, and integrative, which undergirded their assertion that object-orientation is a TC.

In their discussion of TCs as related to Computer Science, Rountree and Rountree (2009, p. 140) suggested *recursion* as a potential example of a TC in that it meets, at least, the criteria of being troublesome, irreversible, and integrative. They referenced a multi-national (UK, US, Sweden), multi-institutional approach consisting of a series of projects from 2006 to 2008 seeking suggestions for TCs in Computing. Although there was no universal consensus, *polymorphism* was a popular candidate TC also (Rountree & Rountree, 2009, p. 143).

State as a TC is proposed by Shinners-Kennedy (2008) saying that state, as he defines it, is “always realized or represented by a value or collection of values”. He argued that the very “everydayness” and pervasiveness of the concept of state (“everyday processes have state”) make it integrative and troublesome, in that changes of state are happening around us that we do not address consciously. By not thinking of state consciously, learners are ill prepared to deal explicitly with state in computer programming.

More recently, K. Sanders et al. (2012, p. 28) indicated that *user-centered design*, the concept that “software is designed for other people to use”, is a good example of a TC in CS. User-centered design, according to Abras, Maloney-Krichmar, and Preece (2004) is defined as “a general term for a philosophy and methods which focus on designing for and involving users in the design of computerized systems” (p. 12). They state that the term became widely used after publication of a book co-authored by University of California San Diego researcher Donald Norman *User-Centered System Design: New Perspectives on Human-Computer Interaction* (Norman & Draper, 1986). Stephanidis also indicates this term as used to “characterize the approach” of techniques to “focus on the requirements of end-users, and provide early feedback to design” (Stephanidis, 2001, p. 8).

2.7 Threshold Concepts and Web Development

Web development, as a specialty under the broader umbrella of Information Technology, clearly has overlap with other areas of the Computing Sciences. As taken from the Association for Computing Machinery (ACM) 2012 Computing Classification System, the following apply to Web development:

Software and its engineering (e.g., E-commerce infrastructure)

Information Systems (e.g., database query processing, query languages, database transaction processing; intranets, extranets, blogs)

World Wide Web (e.g., Web applications, online advertising, Web data description languages)

Security and privacy (e.g., browser security, network security, Web application security)

Human computer interaction (e.g., user interface programming, interaction techniques)

Distributed architectures (e.g., cloud computing, client-server), and

Networks (e.g., network structure, network components) (ACM, 2015).

Web development bears additional challenges beyond traditional systems development precisely because of these many overlaps that require Web developers not only to be familiar with them but also often to work directly in these areas, or at least understand their impact on what the Web developer does. These challenges are considerably more evident today than when Web development had its beginnings.

Web development during the "early days" of the Web, beginning in the 1990s, was simplistic. The Web was a medium primarily for static presentation of data using a variety of formats in a "read-only" mode. Beginning around 2003, a trend toward a more dynamic presentation and active user participation in content management marked the phase called Web 2.0 (Wang & Zahadat, 2009). This required new skill sets in the tool bag of the Web developer for dealing with AJAX (Asynchronous JavaScript and XML), CSS (Cascading Style Sheets), RSS (Really Simple Syndication), and Web services such as SOAP (Simple Object Access Protocol).

Work has continued apace toward the Semantic Web, which some have regarded as synonymous with Web 3.0 (Morris, 2011) while others deemed as a set of complementary concepts (Ankolekar, Krötzsch, Tran, & Vrandečić, 2007). According to the World Wide Web Consortium (W3C) web page, "The Semantic Web is a web of data. [It] provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries" (W3C, 2013). Fuchs et al. (2010) defined Web 2.0 and Web 3.0 in the context of the Web being a *techno-social system* “describing and characterizing the social dynamics and information processes that are part of the Internet.” (p. 43) They posited that Web 2.0 was not a technological novelty, as the technologies existed years beforehand, but rather a social novelty. In any case, these changes demonstrate a continuing evolution of entirely new sets of technologies and terminology needed by the Web developer.

As Web development today involves a diverse and seemingly constantly changing collection of technologies with which the developer must be familiar, it requires proper selection of these technologies, in combinations appropriate to the current task, as well as an understanding of the effect each of these may have on the user experience. Also required, although perhaps more subtly, is how these various technologies work together to provide the overall result. This may indicate a distinction similar to that in Economics, where there are discipline TCs and one or more TCs relating to being able to “model” the dynamic and interconnected nature of a completed Web-based application. Indeed, there may exist a TC of being able mentally to connect the “web” of concepts and technologies involved in Web-based application development. Being overwhelmed with changing technologies and wanting to be relevant, both to students planning soon to use these technologies in the workforce as well as to their future employers, the question of identifying fundamental (and likewise, threshold) concepts in this area sometimes is overlooked.

Alston et al. (2015) remarked upon the paucity of empirical research in Web development, with respect to TCs. They noted the work of T. H. Park and Wiedenbeck (2011) regarding help-seeking activities of students via help forums. The characterization of those areas of difficulties, although not mentioning TCs per se, did look for “challenges” in student understanding. Park and Wiedenbeck primarily referenced various skills and some conceptual content common to the field of computer science. After interviewing higher education instructors in the UK, Alston et al. (2015) suggested two candidate TCs for Web development, *basic programming principles* and *decomposition and abstraction*. These two encompass much, if not all, of the challenges given by Park and Wiedenbeck. Alston et al. (2015) also noted the survey work done by Dorn and Guzdial (2010) of professional Web and graphic designers who had little to no formal computer education and yet were writing computer programs. These professionals recognized that tools and languages they used would continue to change relatively rapidly over time and that they needed to have additional conceptual knowledge. Further, Dorn and Guzdial (2010) remarked that struggling with finding, and especially with applying, new knowledge on the Web resulted in a lack of “instructional guidance about the underlying rationale” and that such “explanations are critical in developing expertise.” (p. 28) This clearly places the burden for knowing and addressing fundamental concepts on the instructional staff. Alston et al. (2015) pointed out “a number of difficulties, however, with using ‘fundamental concepts’ as the basis for uncovering ‘Threshold Concepts’ in curriculum design, particularly within Web development.” (pp. 2:2-2:3) Referencing Davies (2003), they point out that

> “one of the main problems with identifying fundamental concepts as a basis for curriculum design is that they ‘divorce understanding from experience of the world’. This could be likened to the concepts of HTML and CSS; they are both fundamental in the development of Web sites and students need to grasp an understanding of both. However, teaching them in separation would be counterintuitive, especially given their intrinsic relationship in the creation of Web pages.” (Alston et al., 2015, p. 2:3)

This underscores the issue of having a plethora of changing technologies in use today for the development of Web sites – to be effective Web developers, students must learn many tools and their interactions, realizing that these eventually will change. Nevertheless, there also must be a focus on concepts which do not change and which are necessary to comprehend, regardless of the particular technology used. That some, and perhaps all, such concepts are in common with Computing Science as a whole is not surprising. There exists the possibility, however, that the specialty of Web development may have some concepts particular to this area of study. Exactly what these concepts are for Web development, whether in common with Computing Science as a whole or not, remains relatively unexplored. Furthermore, as the above studies indicate, of what has been done, all are essentially qualitative studies involving instructors and/or students, with very little directed towards professionals in the field.

2.8 Extending the Research on Threshold Concepts and Web Development

Miller & Connolly indicated that in spite of Web programming being the dominant model, it is underrepresented in the curriculum of many computing programs, with a concomitant decrease in research of problems associated with the teaching of Web development (Miller & Connolly, 2015). Web development technologies continue to change quickly, presenting educators with the difficulty of covering essentials that remain consistent while keeping up with this dynamic aspect. It is possible that this difficulty is exacerbated by an internal desire on the part of undergraduate students today, at least those considered as “Millennials”. Sendall, Ceccucci, and Peslak (2008) stated,

> "today’s undergraduate students learn very differently from the way many current academicians have. Millennials tend to be more pragmatic; i.e., the subject matter must be ‘useful’ to them. Although there are exceptions, most of them are not in college to explore intellectual ideas; they are focused on learning skills to help them achieve whatever short term goals that they perceive will make them employable and competitive in the marketplace.” (pp. 4-5)

This objective may make the discovery of TCs more difficult, as the rush to “learn the tools” may obscure the need to absorb underlying concepts. This also may explain why studies from the instructor perspective, as in Alston et al. (2015), appear initially and perhaps more frequently than those involving learners.

Rountree and Rountree (2009) indicated that validating TCs via the opinion of students is not suitable as “practitioners define the subject; we define the curriculum”, although it is possible that discovering TCs legitimately can consist of a combination of examining the candidate concepts from the viewpoints of both the instructor and the learners. As Eckerdal et al. (2006) indicated, the areas in which one observes misconceptions among learners are suggestive of places to examine for TCs. Cousin (2006) stated, “The first design principle, then, is to explore (ideally with students) what appear to be the threshold concepts in need of mastery.” (p. 5) The learners can more directly express if they have a clear understanding of concepts, but each only can speak for himself or herself. Cousin’s rationale that “it is difficult for teachers to gaze backwards across thresholds, they need to hear what the students’ misunderstandings and uncertainties are in order to sympathetically engage with them" (p. 5) certainly should not be dismissed, but pragmatically, instructors more likely may identify potential TCs by virtue of experiencing where many students have had difficulties with the concepts. Therefore, due to reasons of ease, consistency, historical perspective, and professional experience, it is tempting to examine only the instructors for possible TCs. However, obtaining a more accurate picture of potential TCs requires a more holistic approach. As only limited research into TCs in Web development exists, there is a need for additional research on this topic within this academic domain. Each group, instructors and students, have both positive contributions and potential negative aspects to inquiry. Students have the immediacy of their learning experience, as against the instructor having long passed through the thresholds and perhaps forgetting the transformative nature of the experience. Instructors, on the other hand, lend many positive aspects, such as the irreversibility and integrative aspects, as well as knowing what concepts remain central above the changing technology. This study approaches professional Web developers as possibly combining the positive aspects of both students and instructors, and at the same time mitigating the negative aspects by being relatively recent to their educational experiences while yet having some reasonable amount of field experience. A further rationale for examining those currently in professional practice within the discipline comes from the emphasis in the ACM Model IT curriculum on real-world experience, both the current 2008 model and the 2017 revision. Statements regarding "…the importance of making professional practice a central component of the curriculum" and "… making professional practice an educational priority" made it clear that professional experience was deemed of high value in the 2008 version (Lunt et al., 2008, pp. 43, 44). Similar statements in the 2017 revision, along with the assertion that "IT programs should adopt a curriculum that integrates learning of professional practice …", demonstrate that the emphasis upon the importance of professional experience has not diminished (Sabin et al., 2017, p. 23). These documents emphasize the importance of real-world experiences, typically effected via capstone courses, internships, team-based courses and real-world projects with real clients. Given such an emphasis upon adopting curricula that incorporates professional practice through courses and work experiences, including current practitioners in a study of concepts critical to preparing students for the professional environment seems most appropriate. As this audience is underrepresented in the research, they remain an intriguing and ripe area for inquiry. It thus became the intention of this study to examine this group for perceptions of a selected candidate TC and to determine the effect of educational and experiential backgrounds on their perceptions.

2.8.1 Candidate Threshold Concepts for Web Development

As an attempt to focus on those candidate TCs most applicable to Web development, this study considered four possible candidate TCs of interest: Object-orientation (OO); Separation of Content, Presentation, and Behavior (SCPB); State; and User-centered Design (UCD).

Object-orientation, a conceptual software design and programming model constructed around the use of objects, is often divided into object-oriented programming, class/object distinction, inheritance, polymorphism, etc., as frequent candidate TCs (Moström et al., 2008; K. Sanders & McCartney, 2016). Object-oriented concepts are used extensively in modern computer programming, including both server-side and client-side Web applications. Regardless of the programming tool used, whether traditional or scripting language, object-oriented concepts are pervasive throughout Web development. While the choice of OOP versus OO may be an issue of granularity or of threshold skills versus TC, as discussed earlier, this study considers the more general OO for greater applicability (encompassing all OO concepts and applications, such as Object-oriented design as well as Object-oriented programming) and because of the later research of Eckerdal et al. (2006) specifically choosing OO, even though others had listed OOP in earlier research. This study takes object-orientation as a TC candidate of interest.

Alston et al. (2015) listed “decomposition and abstraction” as a candidate TC, which they presented as encompassing “separation of content, presentation, and behavior” (SCPB). However, K. Sanders and McCartney (2016) mentioned this separation as specific to web design students. This study considers SCPB as a more specific candidate TC, with respect to Web development, noting that it is often expressed as the model-view-controller design pattern for programming. Therefore, SCPB is also is a candidate of interest for this study.

State, as defined and proposed by Shinners-Kennedy (2008) as a TC, made for a good argument that it is integrative and troublesome and that learners are not prepared to deal with it explicitly. State can be considered as the condition of an entity (e.g., program, object) at a point in time, often the content of the program variables, with any change indicating a change in state. As such, state is a dynamic construct. This seems to be especially applicable to Web development, given the dynamic nature of the Web. Although not specifically mentioned for Web development, understanding stateless Web protocols versus stateful protocols implies a need for Web developers to internalize the concept of state perhaps even more than non-Web developers do. Therefore, State is another candidate of interest.

Finally, user-centered design (UCD) seems to be particularly an issue for Web developers. Although Alston et al. (2015) identified “interface and content design” as an area perceived by instructors as difficult for students, they did not see this as a potential TC. This may be due to viewing the process simply as the envisioning of Web page construction, for which they suggested designing on paper via wireframes, site maps, and the like. However, K. Sanders et al. (2012) proposed that user-centered design, that is, the design of Web pages for the usability of other people, is indeed a good example of a TC and suggested that it met the transformative, integrative, and irreversible criteria. Focusing on end-users' requirements to obtain early feedback to design may prove to be more problematic when facing an amorphous worldwide set of users. This, as well as the general applicability to any software development, made UCD the final choice for a candidate TC of interest.

2.9 Summary

This chapter presented the Threshold Concepts Framework as the theoretical basis for the study. Beginning in 2001 as a research interest in higher education in the United Kingdom, TCs are a relatively new area of study within a variety of academic disciplines. The origins of TCs were examined, when and how they developed, as well as the foundational aspects rooted in conceptual change, connections with social learning theory (communities of practice), and a relationship to Variation Theory.

The dimensions of TCs, their principal characteristics and relationships between one another were presented as were critiques of the theory, which admittedly are few, and rebuttals of said critiques. This chapter examined the identification of various TCs in other academic fields, and then in Computer Science, and lastly looked at research concerning TCs in Web development and proposed several candidate concepts of interest. Noting the issues, both positive and negative, with using instructors or students as sources, this study proposes an extension of that work by utilizing working professionals in Web development. This is intended to mitigate noted weaknesses in the use of students and those in the use of instructors, while retaining the strengths of both sources. This study investigates the perceptions of Web professionals with respect to a given candidate TC and determines how educational background and experiential background affect those perceptions.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Purpose

The purpose for beginning this quantitative study is that all previous research on TCs has been qualitative and essentially no research exists regarding the perception of TCs by professional Web developers. The development of a theory typically utilizes qualitative research and takes an exploratory approach. Quantitative research then brings rigor in testing the exploratory model using a confirmatory approach (Joseph F. Hair, Black, Babin, & Anderson, 2010). Large-scale survey research allows for statements to provide observed indicators for latent (unobservable) constructs and then subsequent rigorous testing of the exploratory model. Therefore, survey research design was chosen to test the theoretical framework.

One objective of this study is to develop a valid and reliable new instrument to measure the perception of candidate TCs by Web development professionals. The final research instrument was distributed via Qualtrics software to a random sampling of Web developers in the U.S. and the data analyzed for the perceptions and the influence, if any, of education and experience.

The process of developing new measures requires several steps to determine appropriate content for the instrument. This chapter describes the development and testing of that survey instrument.

3.2 Instrumentation/Measurement Development and Pilot Study

A new survey instrument was proposed. The final quantitative survey instrument consists of statements designed to elicit opinions of participants on a selected candidate TC and how they perceive the TC, with respect to the defining criteria of a TC. Participants responded to an online questionnaire, which asked some demographic questions, provided a selected candidate TC, gave a simple definition of the listed candidate TC with respect to Web development, and posed a series of Likert scale statements regarding aspects of TCs. The statements for the candidate TC generally are not meant to be applicable only to the selected candidate TC but also germane for use with the other candidate TCs in Web development.

In the process of developing a new survey instrument, particularly where previously validated instruments are not readily applicable, we need to go through several steps to ensure that the items in the instrument are measuring what is intended and whether the instrument is consistent in doing so. We begin by devising items that address the desired outcomes in an operational sense. (Vockell, 1983)

3.2.1 Operational Definitions of Outcome Variables (Constructs)

When we collect data to determine whether a desired outcome is occurring or not, we wish to use a measure for the outcome, aka, the outcome variable (construct). In this case, the constructs are given from the theory and do not have to be developed. They are the particular characteristics of TCs at interest (transformative, irreversible, integrative, troublesome, and bounded.) The first step, therefore, is to devise an operational definition of the variable; the observable events we record as collected data. Learning, however, involves an internal change and cannot be observed directly, meaning that our constructs are *latent* constructs. Consequently, we collect observable evidence that we are willing to accept as an indication that an internal change (learning) has occurred. Such observable evidence would be behavior on which two or more observers would agree that the outcome of interest is occurring. This evidence is the operational definition of an outcome variable, which become the statements used in the survey for the candidate TC. For example, to determine if the outcome for a candidate TC is transformative, meaning that a qualitatively different view of subject matter has occurred, an operational definition might be, “Upon learning about Object-orientation, my ideas about how to program changed significantly.” Please note that the particular TC is not at issue, only whether the statement is a valid measure of the given characteristic of any candidate TC. As this is a possible measure of an internal change, it is important to state multiple observable measures, and as precisely as possible.

Development of the operational definitions for our outcome variables (constructs) is based on a review of the literature, looking for any statements that might be appropriate measures, or using appropriate existing instruments, if any. (In this case, there were no existing measures.) New items are created where no previous operationalization exists. The input of subject matter experts is used to review created statements, comment for validity, and to suggest additional or replacement statements. An additional step of a pilot study using the Q-Sort method provides further support for construct validity and for reliability (Li, 2006; Nahm, Rao, Solis-Galvan, & Ragu-Nathan, 2002; Qrunfleh & Tarafdar, 2014).

Additional guidance in scale development is provided in the measurement scale development framework of Malhotra and Grover (1998), as shown in Figure 6. In the process of developing the scales and subsequent testing, this study followed that outline, modified, of course, as the construct domain is a given.

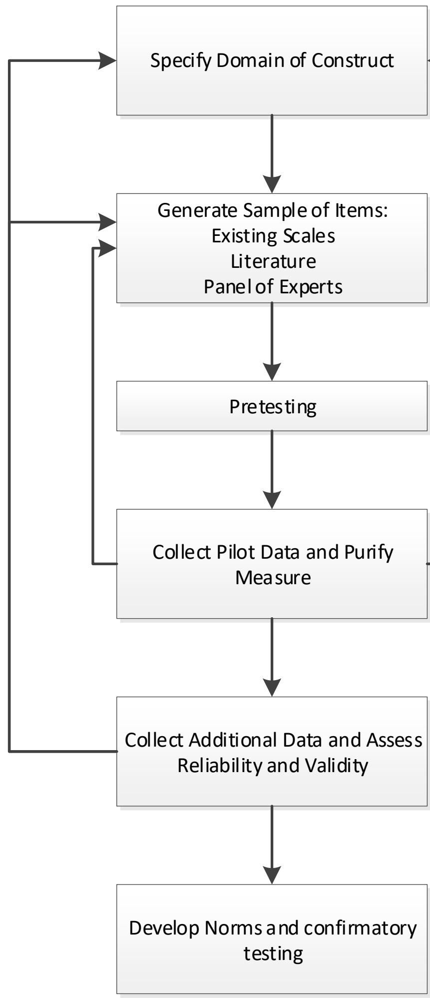

Figure 6. Measurement Scale Development Framework
Malhotra and Grover (1998, p. 412); used by permission

3.2.2 Validity of the Survey Instrument

Validity of an instrument is the extent to which the instrument measures what it is intended to measure. A first step to determining validity is to ascertain whether there is a match between the operational definitions (measures) and the construct(s) they are purported to measure. Normally, this is done by having several persons examine the instrument and make a determination as to the match between the construct definition and the measures. This subjective review, called *face validity*, sometimes is used interchangeably with content validity, but some argue that they should be separate. Nevo (1985) indicates that the use of experts should be avoided for face validity, as their opinions lend to content validity. It is at least somewhat accepted that an instrument having face validity should *appear* valid to respondents (what they infer with respect to what is undergoing measurement) in addition to actually *being* valid (Netemeyer, Bearden, & Sharma, 2003).

Beyond the subjective face validity, there are three main types of validity tests, namely: content validity, criterion validity, and construct validity (Vockell, 1983). Content validity seeks to answer the question of whether the current test covers a representative sample of the subject matter; that is, all relevant items needed to answer the research question. Criterion validity is the degree of correlation between the current test to another measure of performance, ostensibly the predetermined standard. Construct validity is the extent to which the test actually measures what the theory claims; that is, the degree to which the test can be interpreted in terms of the underlying constructs (the psychological quality that we assume explains observed behavior.) "Construct validity is the ultimate goal in the development of an assessment instrument and encompasses all evidence bearing on a measure." (Netemeyer et al., 2003, p. 11) The following summarizes the applicability of each of these to the current study:

(i) Face Validity. A small group of individuals from academia and industry were used to determine whether the instrument appeared to measure what it intends to measure. This is a subjective view of the respondents to the appropriateness of the measures. Items were modified and new ones added subsequent to this review.

(ii) Criterion Validity. Often divided into concurrent and predictive validity. Concurrent validity compares the measure to another outcome assessed (using a different tool) at the same time. An example might be a comparison of CLEP exam scores for college calculus with grades in college calculus to see how the two are related. A predictive validity example, on the other hand, might compare SAT scores with the grade point average of first semester college students to determine the degree to which the SAT is predictive of performance in college. As there are no existing quantitative tests for this subject already held to be valid (the criterion), this test was not used.

(iii) Content Validity. Subject matter experts from academia were presented with the modified list of statements grouped by construct and were asked whether the statement was clear as written, and if not, how it should be corrected; if it applied to the current construct and if not, to which other construct; and for any comments on appropriateness of the statement to the construct. Their feedback resulted in modification of some items to improve clarity.

(iv) Construct Validity. There are two subtypes of construct validity: (a) convergent validity (a measure is associated with things with which it should be, such as alternative measures of the same construct; the extent to which two measures are in fact related) and (b) discriminant validity (a measure is not associated with that with which it should not be; two measures are in reality unrelated when they are not supposed to be related; the extent to which they measure different things.) A convergent construct validity exists when what is expected to be correlated indeed turns out to be correlated; that is, either a positive or a negative correlation is shown. Hence, the null hypothesis, where the correlation coefficient is 0 (H_0: r = 0) and an alternative hypothesis, where the correlation coefficient is not 0 (H_A: r not equal to 0). Such a result shows that H_0 is incorrect and, thus, is rejected. Whereas, in discriminant validity, r = 0; H_0 cannot be rejected. The study used the Q-Sort method to assess for initial convergent and discriminant validity via inter-judge raw agreement, prior to large data collection. Additional tests were applied to ensure validity of the items after the large-scale data collection was completed.

3.2.3 Reliability of the Survey Instrument

Reliability addresses the issue of whether an instrument is consistent in what it measures; that is, the extent to which the results are the same every time. The Q-Sort method was used to assess construct validity and reliability prior to the large data collection. With this Q-Sort, three different measures are used to assess reliability: Moore and Benbasat's item placement ratio, or "hit ratio", (Moore & Benbasat, 1991), Cohen's Kappa (κ) (Cohen, 1960), and Perrault and Leigh's index of reliability (I_r) (Perreault & Leigh, 1989). Appendix Appendix D. Cohen's Kappa and Appendix Appendix E. Perreault and Leigh's Index of Reliability explain the methodology for each of the last two, respectively. The item placement ratio is described in the discussion on the Q-Sort procedures used in the pilot study.

3.2.4 Pilot Study using the Q-Sort Method

The Q-Sort method is an iterative two-stage process providing an assessment of construct validity and reliability of survey items at a pre-testing stage via measures arising from the level of agreement between judges (Moore & Benbasat, 1991; Nahm et al., 2002). This method utilizes panels of two judges per round, each of which has expert knowledge in the field. The judges each receive some instruction, the set of constructs with short definitions, and a randomized list of statements. Each statement is to be sorted into an appropriate construct or listed as not applicable. In each sorting round, the set of judges is unique, so that all judges encounter the instrument only once. Additionally, the judges work independently of one another.

This Q-Sort method used a Web-based survey instrument. There were two stages for each round. In the first stage, participants saw an initial page, which explained the purpose of the project and presented a definition of TCs, some TC examples from other fields, and what the participant was to do, including mechanics of using the web-based survey and how to contact the author, if needed. On a second page, for reference was a list of the five characteristics (Transformative, Irreversible, Troublesome, Integrative, and Bounded) with short definitions of each, along with the statements to sort. The statements appeared in a randomized order for each participant and they selected via a radio button choice the characteristic applicable to a given statement. For each statement, an additional choice of “Does Not Apply” prevented forcing participants to assign any given statement to one of the five characteristics. In the second stage, results for each round were retrieved and analyzed. Items identified in the first stage as not applicable to the expected characteristic or too ambiguous were reworded, removed, or examined for characteristic applicability. The process was repeated until an acceptable level of agreement was obtained.

For this research, two independent judges per round evaluated a set of statements. Three sorting rounds utilized six computing academics, professors in Computer Science, Information Systems, and Information Technology. The presentation of the measures (statements) purposely omitted any particular candidate TC as it is not relevant at this stage. Only the applicability to the characteristics is in question. This prevents the judges from evaluating the candidate TC as opposed to the measures of its characteristics.

Upon receipt of the completed sorts from both judges in a given round, calculations were performed for the assessment of reliability and validity of the sorted statements. The *inter-judge raw agreement* is the count of items where both judges agreed on the categorization (measure of a characteristic) divided by the total number of items. This measure does not depend on whether the judges agreed with the intended characteristic, but only that they both agreed on a single characteristic for the statement. The *item placement ratio*, however, does depend on whether the judges agreed with the intended characteristic. Counting all items sorted into the targeted characteristic and dividing by the total possible (the product of the number of judges, 2 for a round, and the total number of items) yields this measure. This is an overall measure of both validity and reliability, as a higher percentage of "correct" placement indicates a higher degree of inter-judge agreement (expert agreement), and hence validity. Scales are considered to have a high degree of construct validity when there is a high level of correct placement, and such indicates a high likelihood of good reliability. This is more of a qualitative than a rigorously quantitative procedure, as there exists no established guidelines for "good" levels of placement. However, the resulting matrix can prove useful in highlighting potential problem areas. See Appendix B. Sample Item Placement Ratio Calculation for an illustrative sample. Cohen's Kappa (κ) and Perreault and Leigh's index of reliability are calculated for each round, as described in the appendices.

In the first round, the inter-judge raw agreement score was 0.43, κ was 0.29, l_r was 0.56, and item placement ratio was 0.58 (see Appendix E. Perreault and Leigh's Index of Reliability.) The acceptable threshold of 0.65 (Moore & Benbasat, 1991) was not met in any of these four. This resulted in removal of two items and modification of others.

The modified instrument was entered into round two, where the next two academics were used as judges. The inter-judge raw agreement score was 0.53, κ was 0.39, I_r was 0.66, and item placement ratio was 0.68. Results were improved but two were still below the 0.65 threshold and the others only somewhat above. Investigation resulted in some modifications and rewording for clarification.

The third round utilizing the final academic pair of judges resulted in an inter-judge raw agreement of 0.82, κ was 0.78, I_r was 0.88, and item placement ratio was 0.91. As these results indicated high values for all four of the validity and reliability scores, well exceeding the 0.65 threshold, no additional iteration was required. The scales used in the final survey were drawn from this final round after dropping all other statements where the final judges did not agree, save one. The one retained statement had consistent agreement for all judges in all rounds except for one of the final pair of judges. As all other judges also agreed with the original categorization of the statement, this one disagreement was considered an aberration.

An additional statement was added at the beginning of the survey in order to gauge the level of understanding of the respondents regarding the terminology used. Although a definition of the candidate TC was provided, it was thought to be important to verify that they did indeed understand the concept sufficiently; otherwise, their responses might well be no more than guessing.

3.2.5 Reduction of the Scale Items

The final number of statements deemed usable for measures of the characteristics in the large-scale survey was 32. An immediate issue became obvious if all scale items were to be used for all four of the considered candidate TCs. Presenting respondents with all 32 items for each of the four considered concepts all on the same survey results in a survey of considerable length, comprising 128 evaluative statements, in addition to demographics. This resulting survey was considered too lengthy, as the likelihood of respondents simply skipping many statements as they proceeded through the survey, or answering all items the same without real consideration, is high. This would severely increase nonresponse bias and taint the entire survey results. As the scale items were the same, just to be applied to four different candidate TCs, the decision was made to focus the survey on only one candidate TC, as it would reduce the size of the survey dramatically and as other candidate TCs could be tested in subsequent research using the resulting developed and validated instrument. The issue then was determining which of the four candidate TCs originally considered would be utilized for this survey. After closer examination, the selected candidate TC was Separation of Content, Presentation, and Behavior (or MVC).

One last consideration was given to the potential to explore more than a single candidate TC. It was decided that the initial survey results would be examined after approximately half of the respondents are surveyed. By running an exploratory factor analysis on this first group, we might see how many items do not load to the relevant factors, if any. If enough could be dropped, then the remaining half of the respondents could respond to these measures as applied to two candidate TCs without extending the size of the survey significantly. It would add to the richness of the data set to be able not only to compare the two sets on the same measures for a given TC but also compare one TC to another for the same measures, with respect to the second set. In such a case, one of the previous candidate TCs listed above (Object-orientation, State, or User-centered Design) could be used as the second TC. As the factor loading precluded removal of sufficient measures, and as changing the survey could inhibit proper comparison of the two sets, the remaining respondents were presented with the same survey set as the first, examining only the one TC, SCPB.

The rationale was for selecting the single candidate TC was as follows:

<u>For Object-orientation</u>: as noted previously, Object-oriented Programming is taken as a TC because of the paradigm shift in programming, however, even those who took this stand admitted that it might not be granular enough (Boustedt et al., 2007; Luker, 1994). It may be too general to treat Object-orientation as a TC in and of itself. Therefore, it well may be better to

wait until there is more research on the granularity of this topic, lest the concept introduce misconceptions on the part of respondents. That is, responses could reflect individual respondents' "definitions" what is included in Object-orientation, even if a short definition is provided. OO also was not mentioned as a candidate TC with respect to Web development specifically, but typically in regard to programming per se.

For State: Although an interesting and compelling concept for study, the very fact that Shinners-Kennedy (2008) said that state is often not considered consciously makes it potentially problematic, in a similar fashion as OO. If respondents do not understand the concept, or if there is a likelihood that their understanding is sufficiently diverse, this also may cause wide variations in responses, not due to whether the concept meets the attributes of a TC but simply whether all are considering the same thing. This would, however, tend to work in favor of the troublesome nature of a TC, but for the aforementioned reasons, state was discarded. This concept also was not mentioned specifically with Web development in mind, although just as with OO, it clearly applies.

For User-Centered Design: this concept was mentioned as designing software for other people to use, which Web developers clearly do, but more in the context of a TC in CS and more with the idea of involving users directly, as opposed to possibly surrogate users, given that the potential population of users for a Web site can be quite large and contain a significantly unknown set of users. As the focus is "on requirements of end-users" to "provide early feedback to design" (Stephanidis, 2001, p. 8) and as many of whom simply are not available to the Web developer, this concept, although still considered valuable, was deemed potentially to be seen by Web developers as something they do not practice. In fact, it is suspected that they do so, if only by proxy.

For Separation of Content, Presentation, and Behavior (SCPB): Unlike the previous concepts, this concept was specifically mentioned in regard to Web development (K. Sanders & McCartney, 2016). While the concept, perhaps more recognizable in the formulation of model-view-controller (MVC), certainly is applicable to programming areas outside of Web development, it clearly is a concept which Web developers must know how to implement on a regular basis. For them, there should be no lack of clarity as to what this term represents, for it very much is embedded in what they do. Even the variety of development tools utilized for creating responsive Web-based delivery systems reflect this “division of labor”. Due to the nature and difficulties of teaching beginning programmers the basics, such a division often is not taught until later, and thus could be rather troublesome yet transformative and unforgettable once learned. It clearly spans into other avenues of programing (as the term MVC attests) and would appear to be bounded, at least within the computing field. In many respects, SCPB appears to be an ideal candidate TC for Web development and this is the concept selected for this study.

3.3 Large Data Collection

3.3.1 CITI Training and Institutional Review Board (IRB) Approval

As this quantitative study is utilizing survey research for data collection, and therefore dealing with human subjects, and as the researcher is a faculty member at Purdue University Northwest, it was required that the researcher complete Collaborative Institutional Training Initiative (CITI) training both for DePaul and for Purdue University. It also was required that a request be made to the DePaul Institutional Review Board (IRB) for acceptance as exempt research. Training for both institutions was completed and the IRB request for acceptance as exempt research was submitted and approved, utilizing the revised federal guidelines of 2019.

3.3.2 Data Collection Procedures

Item generation for the researcher-designed instrument was done by examining the literature and developing measures, subsequently sent to academic and industry subject matter experts to check for face and content validity and pre-assessed for validity and reliability via a pilot study utilizing the Q-Sort method. Upon completion of the pilot study, the final instrument was created and housed in Qualtrics.

Qualtrics provided a completely anonymous, randomized sample of the target audience solicited on a national basis from their research panels of Web developers containing respondents highly likely to qualify. They utilized their own protocol as a panel aggregator to contact qualified potential respondents via an e-mail invitation, which provided a link to the survey, informed them that the survey is for research purposes only, and provided the expected time required. The survey was introduced by the required DePaul Information Sheet, which included a screening question at the end verifying that respondents were adult (18 or over), English-speaking, working Web development professionals, in order for them to continue with the survey. In the Information Sheet, respondents were informed of the purpose of the study, that they may skip questions or simply exit the survey at any time, the anonymity and confidentiality of the study, and how to contact the researcher and/or advisor.

Reporting capabilities in Qualtrics allowed the data to be downloaded for analysis using other tools. As the instrument is new, the measures were examined to see how well they measured the latent constructs before the hypotheses could be tested.

3.3.3 Large Scale Data Sources

The targeted sampling frame was drawn from professional Web developers working in the United States. The U.S. Bureau of Labor Statistics (BLS) lists some 127,300 Web developers in this market according to the May 2018 Occupational Employment Statistics data released March 29, 2019 (Statistics, 2018). Their definition of Web developer includes many tasks which familiar to Web engineers, and software developers in general, such as analyze, design, create (code), and modify (maintain) code, in this case specifically for a Web site. This would include those who create Web app software using the http protocol (browser-based applications) but would not necessarily exclude native mobile app developers (those using native app development tools not using http.) Included are “back-end developers”, who develop server-side code, accessed indirectly by the user via a front-end application (e.g., a server program which process database requests and return them to the user); “front-end developers” (doing Web design and creating client-side code using technologies such as HTML, CSS, and JavaScript, downloaded to and executed on the user machine); and “full-stack developers” (incorporating both front-end and back-end tasks.) While the BLS allows for the conversion of digital media into a compatible format, the definition excludes those who create such media (“Multimedia artists and animators”), even for Web usage. The occupational code and definition from the BLS is:

> 15-1134 Web Developers
>
> Design, create, and modify Web sites. Analyze user needs to implement Web site content, graphics, performance, and capacity. May integrate Web sites with other computer applications. May convert written, graphic, audio, and video components to compatible Web formats by using software designed to facilitate the creation of Web and multimedia content. Excludes "Multimedia Artists and Animators" (27-1014). (Statistics, 2018)

As the BLS does not have a category for Web Engineering per se, also included would be those for whom this is a title, as opposed to Web Operations Engineer, which has to do with direct operations and not development.

Such Web developers support online business and social media sites including the design, building, and maintenance of these web sites. As it is impractical for this study to measure the entire domain, the sample group was a convenience sample. The group was Web developers who are professionals working in the field of Web development and who have some amount of work experience and zero or more years of formal, post-secondary education, ostensibly undergraduate but possibly graduate work. They may or may not have had formal education in Web development, per se.

3.3.4 Advantages of Web-based Surveys

There are numerous advantages to using a Web-based survey. It is an advantageous means to reach a large sample in a relatively short time (Wright, 2005). For some potential respondents their primary Internet source is their mobile device, increasing not only their connectivity to the Internet, but also their likelihood of a quicker response (Dillman, Smyth, & Christian, 2014). It also can be less expensive than using a mailed survey, which incurs printing as well as return mail costs, in addition to the time required to address and stuff envelopes (Simsek & Veiga, 2001). Having the original data in a machine-readable format avoids data transcription errors, as well as significantly decreasing the time to obtain usable data (Deutskens, De Ruyter, & Wetzels, 2006). Finally, questions can be mandatory and/or used to control the survey flow, even to skipping part or all of the survey based upon an earlier response. This can shorten completion time and eliminate non-qualified respondents.

3.3.5 Disadvantages of Web-based Surveys

Web surveys incur some disadvantages as well. The survey may need to be “mobile-friendly” in order to minimize nonresponse bias (Dillman et al., 2014). Using Qualtrics software, surveys are presented in both a desktop/laptop and a mobile format, both of which display during survey construction. Additionally, the software is able to help automatically identify possible problems with survey presentation and flow.

A significant concern is the quality of the sampling frame and control of the sampling, such that the researcher obtains an appropriate sample size and that those within the sample respond only once, reducing “subject fraud” (Simsek & Veiga, 2001). Those taking the survey should be a good representation of the population, as much as is possible (Wright, 2005). The use of survey panels may maximize response rates, but one must carefully examine the responses to determine whether they are representative or incomplete (Deutskens et al., 2006). The use of Qualtrics panels provides quality checks on sampling appropriate to the population. It also prevents more than a single response per sample member. The possibility of incomplete responses exists, especially as respondents are allowed to skip questions, if desired. Also, the length of the survey poses a greater risk of arbitrary response or non-response, particularly to items further down in the survey, so we attempted to keep the length as short as possible.

3.4 Statistical Methods

A power analysis using G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) was done to determine the needed sample size to achieve sufficient statistical power. After data collection, data was screened for missing variables and outliers. The research data then was analyzed using several statistical tools, including IBM SPSS and SmartPLS v3.3.3 for PLS-SEM (C. M. Ringle, Wende, & Becker, 2015). As the survey was newly constructed and not from previously validated measures, an Exploratory Factor Analysis (EFA) was performed with the first group. While normally done to determine constructs, in this case, the constructs are given from the theory and the EFA was utilized to determine which of the developed measures actually were usable for a given construct. Subsequently, there was some minor survey modification, as previously discussed, and then the remaining data was collected, screened, and analyzed. Partial least squares structural equation modeling (PLS-SEM), a second generational statistical technique, was used with the second group for a confirmatory approach to test the hypotheses.

3.4.1 Power Analysis and Sample Size

Having an appropriate sample size is important to provide enough statistical power for testing the model. The "10 times rule of thumb" is often cited as providing a rough estimate for determining a minimum sample size (Hair, Hult, Ringle, & Sarstedt, 2017). This rule of thumb indicates a minimum sample size that is the larger of

1) 10 times the largest number of formative indicators used to measure a single construct, or
2) 10 times the largest number of structural paths directed at a particular construct in the structural model.

Hair et al. (2017) indicate, however, that a power analysis based on the part of the model having the largest number of predictors (measures) should be used instead (p. 25).

A power analysis provides the probability of rejecting a false null hypothesis (1 - β, where beta (β) is the probability of not rejecting a false null hypothesis, which is a Type 2 error.) It is determined by a function of sample size, alpha (α) level (the probability of a Type 1 error, rejecting a true null hypothesis), and effect size (Cohen, 1988). Effect size is the number of standard deviations between the null mean and the alternate mean. As effect size increases, the power of a statistical test increases.

Calculating a minimum sample size uses a recommended effect size of 0.02, 0.15, or 0.35 for small, medium, and large effects, respectively (Hair et al., 2017). The commonly used power of 80%, the maximum number of measures for any given construct (8, in this study), and a customary alpha level of 5% (2.5% for a two-tailed test) were used as input to a software tool for statistical power analysis, G*Power 3.1.9.4, (Faul et al., 2009) to validate minimum sample size and required statistical power. As the PLS models are estimated through a series of multiple regressions, the F test for linear multiple regression to estimate a fixed model and R^2 deviation from zero was chosen for the statistical test (W. W. Chin, 1998). Utilizing an a priori analysis, for a medium effect size (0.15) at alpha level of 5% (2.5% for a two-tailed test) and 8 predictors, the minimum sample size was predicted at 109 (for one-tailed; 126 for two-tailed) to reach 80% statistical power, and 160 (for one-tailed; 181 for two-tailed) for a 95% level. Following, a post-hoc analysis was run also using G*Power to determine the power of the statistical analysis. Again, using medium effect size (0.15), alpha level of 5% (2.5% for a two-tailed test), and eight predictors, a sample size of 109 yields a power of 80% and a sample size of 160 yields a power of 95%. For two-tailed, the respective values are 126 for 80% power and 181 for 95%. This establishes that a sample size of at least 109 (for one-tailed; 126 for two-tailed) will yield sufficient statistical results and a sample size of at least 160 (for one-tailed; 181 for two-tailed) will yield strong statistical power.

3.4.2 Nonresponse Bias

Although utilizing the Qualtrics survey panels provides some quality checks, as mentioned earlier, respondents still may not be qualified for the survey as per the stated minimum requirements (i.e., working Web developers aged 18 or older), may submit incomplete surveys, or may obviously “click through” by answering all questions the same in an effort to simply “complete” the survey in a minimal time. This concern resulted in the reduction of the survey size from 128 items to 32, exclusive of demographics. Nonresponse can affect the sample frame by causing a non-representative sample, thus limiting the generalizability of the results of the study (Dillman et al., 2014). Therefore, it is important to examine for nonresponse bias prior to continuing with the data analysis. The two common methods for doing so are the independent t-test and the chi-square test, comparison of means and characteristics of the two groups (early vs. late respondents.) (Armstrong & Overton, 1977) Due to the nominal structure of the variables, the chi-square test was utilized. Nonresponse bias was tested for all comparisons between the two groups using chi-square due to using nominal variables. The results given in Appendix G shown no evidence of significant differences between the two groups, and therefore nonresponse bias was not indicated as a serious concern in this study.

3.4.3 Second-Generation Statistical Techniques for Confirmatory Analysis

Analysis of variance (ANOVA), cluster analysis, multiple regressions, are among those statistical techniques typically called first-generation techniques (Lowry & Gaskin, 2014). These techniques have been widely used by researchers, but for more than twenty years, researchers have been turning to second-generation techniques, as they have helped to deal with limitations of the former (Hair et al., 2017).

A family of statistical models, called structural equation modeling (SEM), are known as second-generation techniques which seek to explain relationships among multiple variables by examining the structure of the interrelationships as expressed in a series of equations (similar to a simultaneous series of multiple regression equations) in a specified structural model. SEM can incorporate latent (unobserved) variables as well as manifest (observed) variables into the analysis (Wong, 2013). SEM has similarity to dependence techniques, such as multiple regression, but with SEM, a construct that was the independent variable in one relationship can also be the dependent variable in another relationship. This gives SEM the ability to estimate all of the relationships/equations simultaneously (Hair, Black, Babin, & Anderson, 2010). SEM not only allows multiple measures to represent constructs, but also addresses measure-specific error, whereas other general linear models do not (Weston & Gore Jr., 2006).

A structural equation model has two sub-models: the inner, or structural, model used to specify the relationships between dependent and independent latent variables, and the outer, or measurement, model specifying the relationships between latent variables and their manifest (observed) indicators (Byrne, 2016; O'Rourke & Hatcher, 2013; Wong, 2013). A latent variable in SEM is either exogenous (a construct that explains other constructs in the model) or endogenous (a construct being explained in the model.) Exogenous constructs are identified by having path arrows pointing out from it and having none pointing to it. An endogenous construct has at least one path leading towards it, representing the effect(s) of other construct(s) (see Figure 7.)

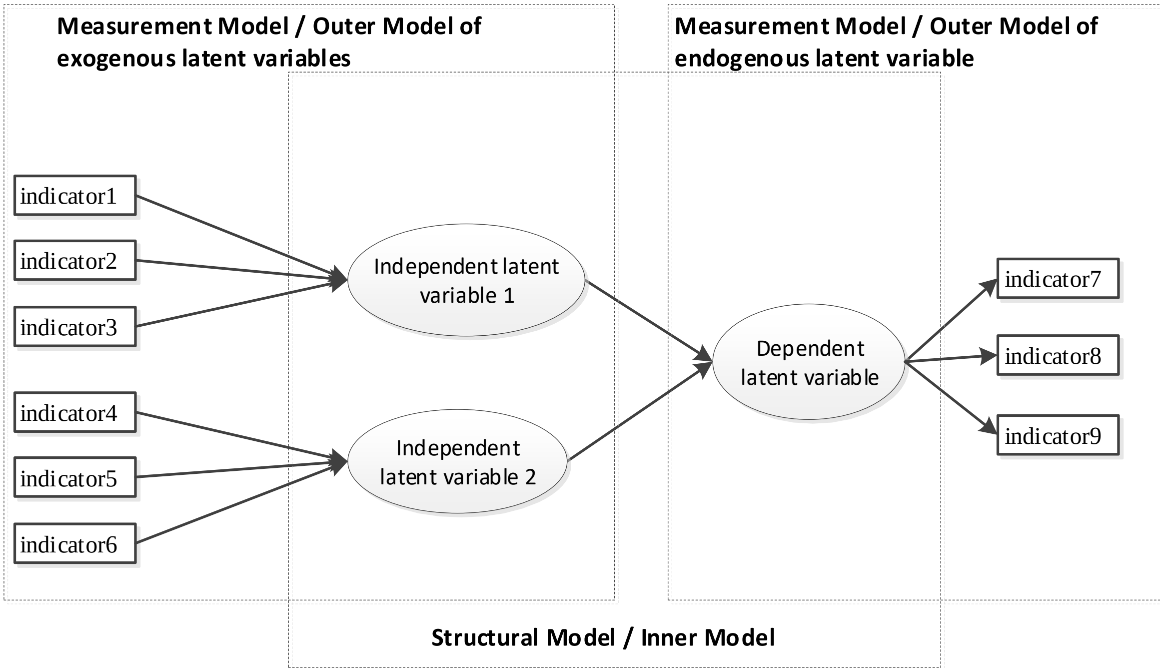

Figure 7. Structural and Measurement Models (Author's image)

SEM has two principal types: covariance-based SEM (CB-SEM) and partial least squares SEM (PLS-SEM), with CB-SEM primarily used for theory confirmation (Byrne, 2016) and PLS-SEM principally used for theory development in exploratory research (Hair et al., 2017). PLS-SEM and CB-SEM are not to be viewed as competitive, but rather as complimentary. Confirmatory work may use either CB-SEM or PLS-SEM (Afthanorhan, 2013; Lowry & Gaskin, 2014). Typically, one will use PLS-SEM when the research goal is predictive of target constructs (Hair et al., 2017; Jöreskog & Wold, 1982; Wong, 2013). As the research primarily is exploratory, dealing with a newly developed survey, all casual effects in the structural model in this study are recursive (unidirectional), and it contains formative constructs, PLS-SEM is used (Hair, Ringle, & Sarstedt, 2011).

CB-SEM develops a theoretical covariance matrix, developed from a set of structural equations, and seeks to minimize the difference with the estimated covariance matrix. CB-SEM requires certain assumptions such as a multivariate normal distribution, no missing data, and a sufficiently large sample size, however, PLS-SEM does not require a normal distribution, can handle smaller (but reasonably so) sample sizes and missing data. PLS-SEM seeks to maximize the explained variance of the latent dependent (endogenous) variable and minimize the error term. Further, when the assumptions of CB-SEM are violated, estimations of the structural model often are more robust with PLS-SEM (Hair et al., 2011).

The models are developed from two theories, measurement theory and structural theory. Structural theory explains how the latent variables, endogenous and exogenous, are related via the paths representative of the relationship between them. There is a left to right sequence for the paths, with variables to the left being the independent variables and any on the right, the dependent variable. The independent variables thus come before the dependent variable and are predictive of it. Note that a given variable may be both a dependent and an independent variable. Exogenous variables are only independent variables, while an endogenous variable may be dependent or both dependent and independent. (Hair et al., 2017) Measurement theory stipulates the measurement of latent constructs (variables) via indicator variables (thus the relationship between latent construct and indicator variables) using either a formative or a reflective measurement model. In a formative model, the directional arrows of the paths point from the indicator(s) to the construct, indicating a predictive relationship in the given direction. In a reflective model, the direction is reversed, from construct to indicator(s), indicating that the construct is casual of the measurement of the indicator(s), or to be more precise, their covariation. PLS-SEM is easily capable of handling both formative and reflective models (Christian M. Ringle, Sarstedt, & Straub, 2012). The reflective approach is widespread with CB-SEM, however, it requires a minimum number of indicators, at least as many unique elements as parameters to be estimated. As for formative models, “authors generally preferred to use PLS-SEM models” (Simonetto, 2012, p. 456).

As with CB-SEM, PLS-SEM tests the proposed model in a two-step process: 1) assessing the reliability and validity of the measurement model, and 2) assessing the structural model.

3.4.4 Measurement Model Evaluation

Assessing the model fit in CFA for CB-SEM uses model fit indices to indicate whether there is a good fit between the hypothesized model and the observed data. There are a number of indices used, and researchers should not rely on a single fit. Commonly reported indices are chi-square (χ^2); Root Mean Square Error of Approximation (RMSEA); Goodness of Fit Index (GFI); Adjusted Goodness of Fit Index (AGFI); Normed-Fit Index (NFI); Non Normed-Fit Index (NNFI); which also is called the Tucker Lewis Index (TLI); Comparative Fit Index (CFI); and, Standardized Root Mean Square Residual (SRMR). (Iacobucci, 2010; Parry, 2019; Schermelleh-Engel, Moosbrugger, & ;Müller, 2003)

Testing the measurement model for PLS-SEM has no recommended goodness of fit measures, although development of such measures is ongoing. CB-SEM goodness-of-fit measures are based upon the differences between the observed (empirical) and theoretical covariance matrices. PLS-SEM maximizes the explained variance (R^2 value) in the path model of the endogenous latent variables. Relying on variances rather than covariances means that covariance-based goodness-of-fit measures are not completely transferable. Additionally, when assessing the measurement model, one must make a distinction between reflective and formative models. Although individual item reliabilities, discriminant validity, and convergent validity can be tested for the individual construct measures of the reflective model, such traditional reliability and validity testing cannot be done for the formative model, as indicators do not highly correlate

(Hair et al., 2011; Simonetto, 2012, p. 456) . However, the significance of coefficients of the formative indicators can be assessed using the bootstrapping procedure, in addition to loadings of the indicators. Indicators with insignificant loading and weight should be dropped from the measurement model.

3.4.4.1 Reflective Measurement Model Evaluation

A reflective measurement model is evaluated via its internal consistency reliability, convergent validity, and discriminant validity.

Measurement reliability is an indicator of the internal consistency in a latent construct. Common tests are Cronbach's alpha (a) (Cronbach, 1951; Kaynak, 2003) and composite reliability, the latter being recommended for PLS-SEM. Composite reliability considers the measurement (outer) loadings with respect to their construct. The reflective loadings should be at 0.70 or above, although values of 0.60 to 0.70 are acceptable for exploratory research (Wynne W. Chin, 1998; Hair et al., 2017; Hair et al., 2011). Values below 0.60 indicate a lack of internal consistency and greater than 0.90 are not desired as they indicate semantically redundant items and not a measure of the construct (all items measuring the same phenomenon of the construct.) The absolute standardized loading of each indicator should be higher than 0.70 as a test of item reliability. Lower loadings could indicate an inappropriate item (negatively impacting content – and construct – validity), a poorly worded item (negatively impacting reliability), or improper transfer of an item from one context to another. The last indicates potential non-generalizability of the item across contexts (Hulland, 1999). However, in general, indicators with loadings from 0.40 and 0.70 should be deleted only if doing so raises the composite reliability above the threshold suggested earlier. The contribution to content validity of the indicator also should be considered prior to removal of the item. However, items with loadings below 0.40 should always be removed (Hair et al., 2011).

Convergent validity is evaluated by indicator reliability and average variance extracted (AVE). Indicator reliability is the size of the outer loading, which should be statistically significant, at a minimum. A rule of thumb is having a standardized outer loading of 0.70 or higher. An AVE of 0.50 is the threshold for acceptance, as this indicates that more than half of the variance of its indicators is explained by the construct.

Discriminant validity can be examined by means of the Fornell-Larcker criterion, cross-loadings, and the newer heterotrait-monotrait (HTMT) ratio of correlations (Hair et al., 2017; Henseler, Ringle, & Sarstedt, 2015). The outer loading of an indicator on its associated construct should be greater than its correlations (cross-loadings) on other constructs in the model. The Fornell-Lacker criterion states that a construct shares more variance with its indicators than with other constructs. Therefore, the square root of the AVE of a construct should be larger than its greatest correlation with any other construct. Stated alternatively, the AVE of the construct should be greater than the squared correlation to any other latent construct (Hair et al., 2017; Hair et al., 2011; Hulland, 1999). The threshold value generally is considered 0.70. However, when all indicator loadings of the constructs vary between 0.60 and 0.80, it is suggested that the Fornell-Lacker criterion performs poorly (Hair et al., 2017). The heterotrait-monotrait (HTMT) ratio of correlations is proposed as a better approach, although cross-loadings and Fornell-Lacker are still considered standards means to assess discriminant validity. HTMT is “the average of the heterotrait-heteromethod (HTHM) correlations (i.e., the correlations of indicators across constructs measuring different phenomena), relative to the average of the monotrait-heteromethod (MTHM) correlations (i.e., the correlations of indicators within the same construct).” (Henseler et al., 2015) Results above 0.85 or 0.90, depending on the similarity of the constructs, are considered to suggest a lack of discriminant validity.

An example serves to clarify the HTMT calculation. Given correlations of all indicators, $x_1 - x_6$ (see Table 1) for the structure shown in Figure 8, where $x_1 - x_3$ are measures for latent construct, L_1 and $x_4 - x_6$ are measures for latent construct L_2, the average MTHM correlations for L_1 is 0.712 and for L_2 is 0.409. The HTHM correlations (shaded area in Table 1) are all pairwise correlations between x_1, x_2, and x_3 and x_4, x_5, and x_6, yielding an average of 0.335. HTMT then is the average HTHM correlations divided by the square root of the HTMT product of the MTHM correlations of L_1 and L_2, as in

$$HTMT(L_1, L_2) = \frac{HTHM_{avg}}{\sqrt{MTHM_{L_1} * MTHM_{L_2}}} = \frac{0.335}{\sqrt{0.712 * 0.409}} = 0.621$$

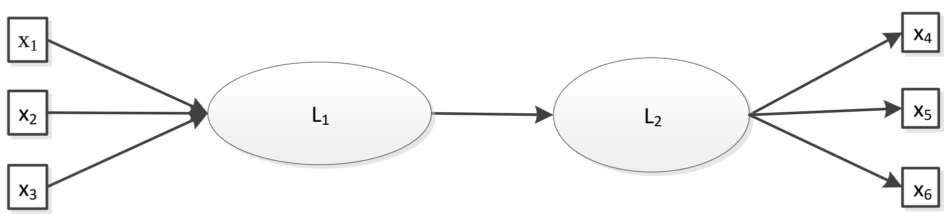

Figure 8. HTMT Example Structural Diagram (Author’s image)

Table 1: Correlations for HTMT Example

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1					
X_2	0.661	1				
X_3	0.703	0.772	1			
X_4	0.339	0.427	0.338	1		
X_5	0.355	0.425	0.398	0.574	1	
X_6	0.242	0.272	0.230	0.368	0.335	1

3.4.4.2 Formative Measurement Model Evaluation

Evaluation of a formative measurement model is done by assessing collinearity between indicators, evaluating the significance and relevance of the outer weights (i.e., of the formative indicators), and bootstrapping.

In formative measurement models, high correlations between indicators (collinearity; when more than two, it is called multicollinearity) can be problematic, may suggest redundant information, causing indicators to be non-significant, as the indicators are not considered interchangeable, unlike reflective indicators. Collinearity is assessed by the variance inflation factor (VIF), which is defined as the reciprocal of the tolerance (TOL). The tolerance is the variance of one formative indicator not explained by others within the same block (for the same construct.) The square root of the VIF represents the degree to which the presence of collinearity has increased the standard error. Therefore, a VIF of 4.00 indicates a doubling of the standard error. VIF values should be less than 5 or a potential collinearity problem is suggested (Hair et al., 2017).

The measured construct should be explained in full by its formative indicators. As the outer weights are standardized, they can be compared with one another. The weights show the relative contribution (importance) of the indicator to forming the construct. The loading of each indicator shows the absolute importance of the indicator. Outer loadings generally should be 0.50 or higher.

As PLS-SEM does not assume normally distributed data, the usual parametric tests of significance are not used. Rather, bootstrapping is used to test the significance of the formative indicators' coefficients. Samples are drawn from the data set with replacement, meaning that as an observation is drawn, it also is returned to the data set before selecting the next observation. The minimum number of bootstrap samples is 5,000 and must at least be as high as the number of valid observations in the data set (Hair et al., 2017). The bootstrap method allows for calculating a Student's t test to see if a given outer weight (w_1) is significantly different from zero (H_0: $w_1 = 0$ and H_1: $w_1 \neq 0$). Using the bootstrap estimated standard error of w_1 ($se^*_{w_1}$) in

$$t = \frac{w_1}{se^*_{w_1}}$$

Critical t-values are 1.65 ($\alpha = 0.10$; two-tailed test) at a 10% significance level, 1.96 ($\alpha = 0.05$; two-tailed) for 5% significance, and 2.57 ($\alpha = 0.01$; two-tailed) for 1% significance.

3.4.5 Structural Model Evaluation

After the reliability and validity of the measurement model is established, the structural model can be analyzed. This involves examining the relationships between constructs and the predictive capability of the model. Recall that CB-SEM estimates parameters to minimize the difference between the theoretical covariances and those of the sample. The various measures of goodness-of-fit are based on this difference between these two covariance matrices. PLS-SEM estimates parameters in order to maximize the explained variance of the endogenous latent variables. PLS-SEM has a different statistical objective and the idea of fit therefore is not completely transferrable. PLS-SEM therefore should employ nonparametric prediction-oriented measures for evaluation of the structural model (W. W. Chin, 1998).

The structural model primarily is assessed based on criteria determined by the model's predictive capability of the endogenous constructs. Key criteria for structural model assessment in PLS-SEM are the R^2 (coefficients of determination) values, predictive relevance (Q^2), size and significance of path coefficients, f^2 effect sizes, and q^2 effect sizes (Hair et al., 2017; Christian M. Ringle et al., 2012). Before assessing these criteria, the structural model must be examined for collinearity. This is done in the same manner as for formative measurement models, as described earlier, but looking at predictor constructs rather than indicators (Hair et al., 2017).

One of the most common measures used is R^2. Values of 0.75, 0.50, or 0.25 are considered as substantial, moderate, or weak (Hair et al., 2011). This indicates that the greater the R^2 value, the better the structural model fit.

The path coefficient are standardized values which usually are between +1 and -1, with values close to +1 indicative of a strong positive relationship, and values close to -1 of a strong negative relationship. A path coefficient equal to 0 is assumed to be insignificant (Hair et al., 2017). Once again, bootstrapping is used to assess the significance of the path coefficients. As with the formative measurement model, a minimum of 5,000 samples (>= the number of observations in the original sample) are needed, and the relationship is considered significant when the calculated t-value exceeds the critical t-value.

The f^2 effect size is the change in R^2 when a particular exogenous construct is removed from the model. This can indicate whether or not the removed construct has a substantial impact on the endogenous construct(s). The f_2 effect size is calculated as

$$f^2 = \frac{R^2_{included} - R^2_{excluded}}{1 - R^2_{included}}$$

where $R^2_{included}$ is the R^2 for the endogenous variable when the selected exogenous variable is included. $R^2_{excluded}$ is the R^2 of the endogenous variable when the selected exogenous variable is excluded.

Stone-Geisser's Q^2 value (Geisser, 1974; Stone, 1974) measures the model's out-of-sample predictive power (predictive relevance.) The Q^2 value is obtained through the blindfolding procedure, which is a sample reuse technique where data points are omitted and then treated as missing values. An omission distance, D, is specified, which causes every *d*th data point to be omitted. While this value is entered as a parameter typically between 5 and 10, it must be selected such that division of the number of observations divided by D does not result in an integer. Estimates are used to predict the missing data points and the difference between the predicted values and the true (omitted) ones is input into the Q^2 measure. Thus the obtained Q^2 values provide a measure of how well the model is able to predict the originally observed values, with values greater than 0 indicating predictive relevance for the given reflective endogenous construct. (Hair et al., 2017)

Similar to the f^2 effect size for R^2, the q^2 effect size is a measure of the relative impact of predictive relevance. The q^2 effect size measure is calculated as

$$q^2 = \frac{Q^2_{included} - Q^2_{excluded}}{1 - Q^2_{included}}$$

Also, like the f^2 calculation, $Q^2_{included}$ is the Stone-Geisser's Q^2 (Q^2) of the endogenous variable, indicating the structural model's predictive relevance when the selected exogenous variable is included. $Q^2_{excluded}$ is the Q^2 of the endogenous variable when the selected exogenous variable is excluded. Again, just as with f^2, the effect for q^2 is assumed to be small if the calculated value is 0.02, medium if the value is 0.15 and large if the value is 0.35 (Hair et al., 2017).

CHAPTER 4: RESULTS OF THE STUDY

This chapter presents the findings of the research. The first part presents the demographic characteristics of respondents in both data sets. The second part is a description of the exploratory factor analysis done using SPSS on the first set (Sample 1) of data collected. The third part describes the results of the PLS-SEM model on the second data set (Sample 2) collected. The final part summarizes the finding of the empirical research.

4.1 Respondent Demographics

Respondents were required to verify that they were an English-speaking adult (18 years old or over) and currently employed as a Web development professional. Respondents provided for birth year (to determine age), gender, and the zip code of their work location, as the survey was sent across the U.S. In Sample 1, which was analyzed using SPSS, the ages of the respondents ranged from 19 years to 66 years old, with an average of 37.5 years. In Sample 2, analyzed using PLS, ages ranged from 19 years to 67 years old, with an average of 38.2 years in age. Table 2 provides these results. Of the 237 in Sample 1, the majority were male at 155 with 82 indicating female, yielding percentages of 65% and 35%, respectively. Of the 248 in Sample 2, the majority were male at 177 with 71 as female, yielding percentages of 71% and 29%, respectively. Table **3** summarizes gender results.

Table 2: Respondent Ages

	Sample 1 (N)	Sample 2 (N)
Min Age	19	19
Max Age	66	67
Average Age	37.5	38.2

Table 3: Respondent Gender

	Sample 1 (N)	Sample 1 (%)	Sample 2 (N)	Sample 2 (%)
Male	155	65%	177	71%
Female	82	35%	71	29%
Total	237		248	

Additionally, respondents were asked about employment characteristics, whether their status was full-time at 40 hours or more/week, part-time, self-employed, or other; primary job function as a text field; years of work experience in computing (in increments); years of Web development experience (in increments); percent of Web development consisting of writing code (in increments); and primary job type, that is type of developer, whether front-end (user-side), back-end (server-side), full-stack (both front and back), or none of these. Table 4 summarizes responses for employment characteristics.

Finally, the demographics concluded with educational background, whether no post-secondary education, some college but no degree, a certificate, an associate's degree, bachelor's degree, master's degree, professional degree (e.g., MD, DDS, JD, etc.), and PhD. Respondents were encouraged to select all that applied, but only the highest degree level was of interest. Fully 25% of Sample 1 have no college with an additional 10% having some college but remaining below a bachelor's degree, while two-thirds hold an undergraduate or graduate degree. Sample 2 has similar results of 24%, 12%, and 63%, respectfully. Table 5 summarizes these results.

Table 4: Employment Characteristics of Respondents

	Sample 1 (N)	Sample 1 (%)	Sample 2 (N)	Sample 2 (%)
Full-time	203	86%	218	88%
Part-time	11	5%	13	5%
Self-employed	19	8%	17	7%
Other	4	2%	0	0%
Computing Experience				
< 1 year	9	4%	11	4%
1 to 2 years	27	11%	26	(round) 10%
3 to 5 years	57	24%	69	28%
6 to 10 years	74	31%	63	25%
> 10 years	70	30%	79	32%
Web Dev Experience				
< 1 year	24	10%	30	12%
1 to 2 years	36	15%	37	15%
3 to 5 years	68	29%	85	34%
6 to 10 years	67	28%	55	22%
> 10 years	42	18%	41	17%
Percent coding				
< 10 %	38	16%	33	13%
10% to 25%	31	13%	30	12%
25% to 50%	59	25%	75	30%
50% to 75%	67	28%	60	24%
> 75%	42	18%	50	20%
Job Type				
Front-stack	68	29%	63	25%
Back-stack	19	8%	30	12%
Full-stack	108	46%	109	44%
None of these	42	18%	46	19%

(all amounts rounded to whole numbers)

Table 5: Educational Background of Respondents

	Sample 1 (N)	Sample 1 (%)	Sample 2 (N)	Sample 2 (%)
None	33	14%	37	15%
Certificate	26	11%	23	9%
Some, no degree	6	3%	12	5%
Associate degree	16	7%	18	7%
BA/BS degree	80	34%	57	23%
MS degree	58	24%	85	34%
Professional Dr.	9	4%	3	1%
PhD	9	4%	13	5%

4.2 Results from the SPSS Analysis

The Sample 1 data collection was used for Exploratory Factor Analysis (EFA) in SPSS, as the survey consisted of a newly created scale. Factor analysis is used to group variables into dimensions identifying latent variables, that is, our constructs. It also is used to reduce the number of variables, thus simplifying the data, and to evaluate construct validity of a scale (Williams, Onsman, & Brown, 2010). While typically used to develop the theoretical constructs, in this case, we are targeting the theoretical constructs provided by previous work, as detailed earlier. An EFA is performed because the scale used is newly developed, we want to reduce the number of variables, and construct validity must be established.

In using SPSS, the default extraction technique is Principal Component Analysis (PCA). From PCA we obtained an initial factor solution with a reduced number of variables, which is a product of factor analysis, as mentioned earlier. No linearity or normality is assumed in PCA, as the process is to determine the relationship of each variable to an underlying factor as expressed by the factor loadings. The factor loadings are determined by the decomposition of the set of measures into orthogonal components, ordered by the amount of variance explained by a factor for each observed variable, that is, the eigenvalue for each. A factor with an eigenvalue >= 1 explains more variance than a single observed variable, and can be used in other analyses, with factors explaining the least amount of variance being discarded. While both normal and non-normal components are guaranteed to be uncorrelated in either case, independence of the components is not guaranteed with non-normality (Kim & Kim, 2012). Given the likelihood of dependence among the constructs presented from the theory, independence was not expected. This lends to why the initial results primarily loaded on a single factor and why exploring extraction over differing sets of observed variables was necessary.

Four factor loadings (components) were identified with eigenvalues >= 1, as shown in Table 6. Varimax rotation was used as it “minimizes the number of variables that have high loadings on each factor and works to make small loadings even smaller” (Yong & Pearce, 2013, p. 84). It is an orthogonal rotation for uncorrelated factors (Costello & Osborne, 2005). The Pearson correlations below .30 were excluded as lower scores suggest variables have a very weak relationship (Yong & Pearce, 2013). The resulting rotated component matrix is shown in Table 7.

Table 6: PCA with Varimax Total Variance Explained

Total Variance Explained									
	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.780	38.899	38.899	7.780	38.899	38.899	3.874	19.372	19.372
2	1.603	8.017	46.916	1.603	8.017	46.916	2.820	14.100	33.472
3	1.115	5.575	52.492	1.115	5.575	52.492	2.701	13.506	46.978
4	1.001	5.005	57.497	1.001	5.005	57.497	2.104	10.519	57.497
5	.837	4.184	61.682						
6	.794	3.969	65.651						
7	.782	3.911	69.562						
8	.761	3.804	73.366						
9	.651	3.255	76.621						
10	.590	2.951	79.573						
11	.549	2.746	82.318						
12	.525	2.626	84.944						
13	.500	2.498	87.442						
14	.480	2.401	89.843						
15	.444	2.221	92.064						
16	.387	1.934	93.998						
17	.361	1.804	95.803						
18	.309	1.545	97.348						
19	.270	1.350	98.698						
20	.260	1.302	100.000						
Extraction Method: Principal Component Analysis.									

Table 7: PCA Rotated Component Matrix

Rotated Component Matrix[a]				
	Component			
	1	2	3	4
Integ02	.750			
Trans04	.732			
Trans03	.660			
Trans07	.598	.364		
Trans02	.584			.527
Integ05	.567			.307
Integ08	.490	.309		
Trans06	.489	.465		
Irrev06		.706		
Irrev04	.366	.683		
Irrev03		.642		
Irrev05		.579		.516
Irrev02	.426	.444		
Trbl03			.809	
Trbl05			.772	
Trbl02			.717	
Trbl04	.453		.628	
Boun01				.644
Boun02				.608
Boun05	.341	.432		.448

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.[a]

a. Rotation converged in 7 iterations.

Items identified as Troublesome clearly loaded as a factor, with only Trbl04 loading on more than one. Items for Irreversible loaded principally on a single factor, with the Irrev02 and Irrev05 only slightly favoring factor 2, and Irrev04 having some mild loading onto factor 1. Of the three Bounded variables, Bound05 loaded across three factors, with the greatest being factor 4, where the other Bounded items loaded. Transformative and Integrative items loaded together as a single factor. This is not overly surprising, considering that if a concept transforms one's perception of the field, it well may help to substantially integrate subject knowledge. As the Transformative items numbered more than those for Integrative, and as their sizes are as large or larger, Transformative was the loading retained for analysis.

Given the results above, we continued with the selected variables for use with the Confirmatory Factor Analysis using PLS-SEM for model development and evaluation of the Sample 2 data collection.

4.3 Results from the PLS-SEM Analysis

A second set of respondent group data, Sample 2, was collected for use with PLS-SEM. Although the Q-Sort provided a measure of validity for the measurement scale, we indicated that further tests would be used after completion of the large-scale data collection. Appropriate methods are based upon the statistical method used, and for PLS-SEM, the various approaches recommended are AVE, composite reliability, assessing the outer loadings, Fornell and Larcker's internal consistency measure, and assessment of cross factor loadings (Hair et al., 2017).

One salient point here is that as the research progressed, it became clear that the model used was a more complex variant called a hierarchical component model (HCM), containing two levels of constructs. The sub-dimensions, or Lower Order Constructs (LOCs), of Troublesome, Transformative, Irreversible, and Bounded, capture the more abstract Higher Order Construct (HOC) of Concept as a TC. This model is a reflective-reflective HCM, with the LOCs in a reflective relationship with the HOC, and the LOCs measured by reflective indicators. Each construct is required by PLS-SEM to have at least one indicator. To represent the measurement model for the HOC, which is itself an abstract representation of the LOCs, a repeated indicators approach was employed as is typical, with the full set of indicators used in the LOCs also assigned to the HOC (Hair et al., 2017, p. 283). This is indicated by the '+' sign in the HOC (see Figure 9.)

4.3.1 Assessing Reliability and Validity – Measurement Model Evaluation

The measurement model, developed consistent with the literature, utilized constructs in a reflective measurement model, as in each case, the measures were caused by the construct, rather than causing the construct, and were highly correlated with one another. Prior to testing the hypotheses, evaluating the reliability and validity of the measures and constructs is needed.

SmartPLS 3.3.3 was acquired from its website (www.smartpls.de) and utilized for the PLS-SEM analysis.

4.3.1.1 Reliability and Validity – LOC Measurement Model Evaluation

Measures for internal consistency reliability traditionally have included Cronbach's alpha. Given that the test is sensitive to number of measures and "generally tends to underestimate the internal consistency reliability", it is suggested that composite reliability is more appropriate, but that the reporting of both should be included (Hair et al., 2017, p. 111) As shown in Table 8, all scores, save those for Bounded, were above the 0.70 threshold. Bounded was slightly low only for Cronbach's alpha and its composite reliability was well within range. Therefore, there is no reliability issue.

Table 8: Measurement Reliability Scores of Constructs

Constructs	Composite Reliability	Cronbach's Alpha
Transformative	0.886	0.839
Troublesome	0.893	0.840
Irreversible	0.902	0.864
Bounded	0.814	0.664

Two measures for convergent validity are the outer loadings of the indicators, the sizes of which commonly are called indicator reliability, and AVE. The outer loading is the result of regression of a measure on its corresponding latent construct. The earlier explanation in § 3.4.4.1 referencing a threshold of 0.70 is a normal lower limit, however lower results often are obtained when a study uses a newly developed scale, which this study does (Hair et al., 2017, p. 113). Further, we noted earlier that indicators with loadings below 0.70 should only be removed if it raises the composite reliability above the earlier threshold, with the understanding that any below 0.40 should always be dropped. Only the Trans07 outer loading is below the 0.70 threshold, and only barely so at 0.695. All were statistically significant. The outer loadings of the measurement items are shown in Table 9 and being within the above limits, they confirm the convergent validity of the constructs.

Table 9: Outer Loadings of Measurement Items

Items	Bounded	Irreversible	Transformative	Troublesome
Boun01	0.759			
Boun02	0.729			
Boun05	0.823			
Irrev02		0.769		
Irrev03		0.762		
Irrev04		0.820		
Irrev05		0.846		
Irrev06		0.828		
Trans02			0.782	
Trans03			0.790	
Trans04			0.818	
Trans06			0.812	
Trans07			0.695	
Trbl02				0.807
Trbl03				0.837
Trbl04				0.820
Trbl05				0.823

The AVE, as mentioned in § 3.4, indicates the degree to which the variance of the indicators is explained by the latent construct. Being above the threshold of 0.50 is a confirmation of the convergent validity of all constructs, as shown in Table 10.

Table 10: Average Variance Extracted of Constructs

Construct	Average Variance Extracted
Transformative	0.609
Troublesome	0.675
Irreversible	0.649
Bounded	0.595

With the model being a reflective-reflective hierarchical component model (HCM) as described, the HOC, being an abstract representation of the LOCs, contains within it the LOCs, incorporates all of the indicators of all of the LOCs, and thus all are measuring the same thing (the concept as a TC). Additionally, having an interconnectedness, as described by Davies and Mangan (2005) and discussed in § 2.3, discriminant validity may not be well established. Therefore, as Hair et al. (2017) suggest, the discriminant validity between the HOC and the LOCs, as well as within the LOCs were ignored because of the high correlations among these constructs.

With these results, we conclude that the LOC measurement model is both valid and reliable. Therefore, we can state that the indicators (measures) are components of the LOCs, meaning that the observable measures do measure their respective latent variables in the model.

4.3.1.2 Reliability and Validity – HOC Measurement Model Evaluation

Assessment of the measurement model of the HOC differs from that of the other (LOC) constructs, in that it is concerned with the relationships of the HOC and its LOCs, and not with the repeated indicator variables . These relationships correspond to loadings, but are mapped as path coefficients in PLS-SEM. In a reflective-reflective HCM, these are used as input to estimate AVE, and to calculate composite reliability and Cronbach's alpha manually by using the formulas as provided by Hair et al. (2018) and shown in Appendix H and Appendix I, respectively.

The loadings (path coefficients) for each construct exceed the 0.70 normal threshold for size, as a confirmation of convergent validity. Using the loadings as input to estimate the AVE of the HOC (concept as a TC), the mean of the HCM's squared loadings is calculated to be 0.805, as shown in Table 11. Being above the threshold of 0.50 is also a confirmation of the convergent validity of the HOC. However, both results depend on significance testing via bootstrapping.

Table 11: HCM Loadings and Squared Loadings

Construct	Loading	Squared Loading
Bounded	0.885	0.783
Irreversible	0.913	0.834
Transformative	0.928	0.861
Troublesome	0.861	0.741
Sum	3.587	3.219
Mean	0.897	0.805

Using the calculations for composite reliability (CR) yields

CR = (0.928+0.861+0.913+0.885)^2 /

((0.928+0.861+0.913+0.885)^2 +

(1-0.928^2) + (1-0.861^2) + (1-0.913^2) + (1-0.885^2))

= 0.943

With four LOCs with an average correlation of 0.897 and M = 4, the calculation for the (standardized) Cronbach's alpha of the HCM yields:

Cronbach's alpha = 4*0.897 / (1 + (4-1) * 0.897)

= 3.588 / (1 + 2.691)

= 3.588 / 3.691

= 0.972

As both Cronbach's alpha and composite reliability are well above the 0.70 threshold, there is no reliability issue. As mentioned earlier, discriminant validity is ignored.

With no validity or reliability issues identified, we can say, assuming that the loadings in the HOC Measurement Model are significant, this indicates that the LOCs are explained by the more general HOC, being more concrete components of the HOC. Therefore, the characteristics of Transformative, Troublesome, Irreversible, and Bounded are components of the TC.

4.3.2 Confirmatory Testing – The Structural Model

After confirming the reliability and construct validity of the measurement model for both the LOCs and the HOC to be acceptable, the structural model must be checked for collinearity before interpreting the results. Collinearity is assessed by the variance inflation factors (VIFs), which are calculated by the SmartPLS software. As there is only a single predictor construct, the HOC, for each of the LOCs, they all yield a VIF of 1.000. Being less than the recommended threshold of five (Hair et al., 2017, p. 145), there is no issue with collinearity.

After confirming no collinearity in the structural model, the significance of the path coefficients (i.e., the relationships) was evaluated by running the bootstrapping resampling method, using 5,000 samples from the 248 cases in the original sample, for a two-tailed test at 0.05 confidence level. This was done to determine whether the effect of the HOC on each LOC is significant, with the strength of each relationship shown by of the path coefficients. The results indicate confidence that the LOC being dimensions of the HOC is warranted. The resulting t-values for all path coefficients in the structural model are shown in Table 12.

Table 12: Bootstrapping Analysis Results of Path Coefficients of LOCs

Paths	Original Sample	Sample Mean	Standard Error	t Values
Concept as a TC -> Bounded	0.885	0.885	0.015	57.345***
Concept as a TC -> Irreversible	0.913	0.912	0.016	56.903***
Concept as a TC -> Transformative	0.928	0.927	0.015	63.123***
Concept as a TC -> Troublesome	0.861	0.862	0.021	40.707***

***$p < 0.01$

As a measure of the model's out-of-sample predictive power, the Q^2 values obtained through the blindfolding procedure for each LOC are shown in Table **13**. These results were obtained by using the SmartPLS default value of 7 for the omission distance (D). As all are well above 0, indicating substantial predictive relevance.

Table 13: Q^2 values of Endogenous Constructs

Construct	Bounded	Irreversible	Transformative	Troublesome
Concept as a TC	0.446	0.501	0.504	0.491

Adjusted R^2 values were examined as an evaluation of the explained variance of each endogenous (LOCs) by the exogenous variable having a path to it (the HOC.) In this case, there is only a single exogenous variable, having paths to each of the exogenous variables. R^2 values of 0.25 are considered weak, 0.50 moderate, and 0.75 substantial (Hair et al., 2017). Only Troublesome was marginally below the 0.75 limit for substantial, with the others all above. The endogenous variables and their respective R^2 values are presented in Table 14.

Table 14: Adjusted R^2 Values of Endogenous Variables

Endogenous Variables	R^2 Adjusted
Bounded	0.782
Irreversible	0.832
Transformative	0.861
Troublesome	0.741

To assess the relative importance of an exogenous variable as a predictor of a related endogenous variable, the effect sizes of significant path coefficients are used. The first method is to assess f^2. While f^2 is calculated by using $R^2_{included}$ and $R^2_{excluded}$, as explained in § 3.4.5, SmartPLS 3.3.3 calculates these, as illustrated in Table 15.

Table 15: f^2 Effect Sizes of Exogenous Variables

Construct	Bounded	Irreversible	Transformative	Troublesome
Concept as a TC	3.597	4.988	6.234	2.875

In this case, there is only one exogenous variable and four related endogenous variables. All are well above the threshold of 0.35 for large effect (Hair et al., 2017) indicating that the effect of the exogenous variable (HOC) on each of the endogenous variables (LOCs) is large. Thus, the exogenous variable is an important predictor of the endogenous variables.

A second method for testing effect size uses q^2. SmartPLS 3.3.3 does not calculate these values, however, they can be manually calculated using Q^2, via the formula presented in § 3.4.5. However, as that formula requires removal of one exogenous variable at a time, and we have only one, this calculation cannot be done.

As the results of both the Measurement Model evaluations and the Structural Model evaluation show, Web developers did perceive the selected concept, the Separation of Content, Presentation, and Behavior, as a TC, by positive and significant responses to the measures for the Troublesome, Transformative, Irreversible, and Bounded characteristics of a TC. This addresses the first research question.

4.3.3 Evaluation of Demographic Indicators

To determine whether the specific demographics of education and experience impacted the perception of the respondents (the remaining two research questions), the results of the bootstrapping technique were examined, with respect to these items. As age can be related to all the initial demographic indicators selected, with the exception of a degree in the computing field, it was included in the model. The bootstrapping results are shown in Table 16.

Table 16: Bootstrapping Analysis Results of Demographic Indicators

Paths	Original Sample	Sample Mean	Standard Error	t Values
EdYrs-> Concept as a TC	-0.099	-0.102	0.112	0.882
CptDegr -> Concept as a TC	-0.124	-0.126	0.062	2.014**
EdLvl -> Concept as a TC	0.216	0.215	0.093	2.329**
WorkExp -> Concept as a TC	-0.294	-0.287	0.089	3.287***
WebExp-> Concept as a TC	0.429	0.423	0.091	4.719***
Age -> Concept as a TC	0.177	0.181	0.085	2.071**

*p<0.10, **p<0.05, *** p< 0.01

Of the six items examined, only the years since last in formal education (EdYrs) was shown as not significant. All others were either significant or very significant, some with a positive relationship and some negative. Of those included in our hypotheses, two had positive relationships and three had negative relationships. Age, while not included as a hypothesis, had a positive relationship. While these results address the hypotheses presented, additional analysis was performed to determine where differences exist in those items of statistical significance.

To determine a single measure from the 1 to 5 Likert scale for the respondent perceiving the concept as a TC, the mean of all used indicators for the characteristics of Troublesome, Transformative, Irreversible, and Bounded was calculated for each respondent, as using a mean or median is a common practice (Sullivan & Artino Jr, 2013).

An independent samples t-test was conducted using SPSS to compare the perception of the TC in those with a degree in a computing field and those without a computing degree (CptDegr), whether degreed or not. There was not a significant difference in the scores for those with ($M=3.9$, $SD=0.72$) and those without ($M=4.033$, $SD=0.805$) a computing degree; $t(246) = 1.33$, $p = 0.185$.

A one-way ANOVA was done for each significant demographic indicator having more than two groupings, and when those results were significant, a post hoc analyses using Fischer's Least Significant Difference (LSD) post hoc test followed to determine the groups where specific differences occurred. All tests were completed with SPSS and using an alpha level of .05.

For Education Level (EdLvl), responses were grouped as 1) No college/certificate only, 2) at least some undergraduate, and 3) graduate degree. A one-way ANOVA showed a significant difference at the $p<.05$ level, [$F(2, 245) = 12.555$, $p = .000$]. Post hoc comparisons using Fischer's LSD test suggest a significant difference between the scores for those with a graduate degree (M=4.256, SD=.628) and those with only undergraduate work (M=3.731, SD=.775), p =.000 or no post-secondary education or only certificate work (M=3.875, SD=.853), p =.002. Comparison between no post-secondary education or only certificate work and any undergraduate work was not significant.

For Computing Work Experience in Years, results were grouped as < 1 year, 1-2 years, 3-5 years, 6-10 years, and >10 years. The result of a one-way ANOVA at the p<.05 level was not significant, [F(4, 243) = .940, p = .441].

For Web development Experience in Years, results were grouped as < 1 year, 1-2 years, 3-5 years, 6-10 years, and >10 years, and one-way ANOVA showed a significant difference at the p<.05 level, [F(4, 243) = 9.771, p = .000]. Post hoc comparisons using the Fischer's LSD test suggest a significant difference between the scores for <1 year (M=3.261, SD=1.223) with all other ranges, 1-2 years (M=3.887, SD=.579), p =.001; 3-5 years (M=4.088, SD=.641), p =.000; 6-10 years (M=4.053, SD=.668), p =.000; >10 years (M=4.265, SD=.575), p =.000. There also was a significant difference between 1-2 years (M=3.887, SD=.579) and >10 years (M=4.265, SD=.575), p =.022. There were no significant differences for other pairings.

For Age, responses were grouped as <30, 30-39, 40-49, and 50+ years. A one-way ANOVA showed a significant difference between groups at the p<.05 level, [F(3, 244) = 3.794, p = .011]. Post hoc comparisons using the Fischer's LSD test suggest a significant difference between the <30 (M=3.752, SD=.612) and the 40-49 (M=4.168, SD=.68), p =.002; and the 40-49 (M=4.168, SD=.68) and those 50+ (M=3.830, SD=.773), p =.042. Those in the 30-39 group (M=3.941, SD=.782) were not significant in difference when compared to the others.

4.4 Summary of the Results

In conclusion, the results of the SPSS analysis occasioned a reduction of the number of measures and the removal of the Integrative endogenous variable from the proposed research model, as it did not load as a separate factor. This resulted in a modified research model, as shown in Figure 9.

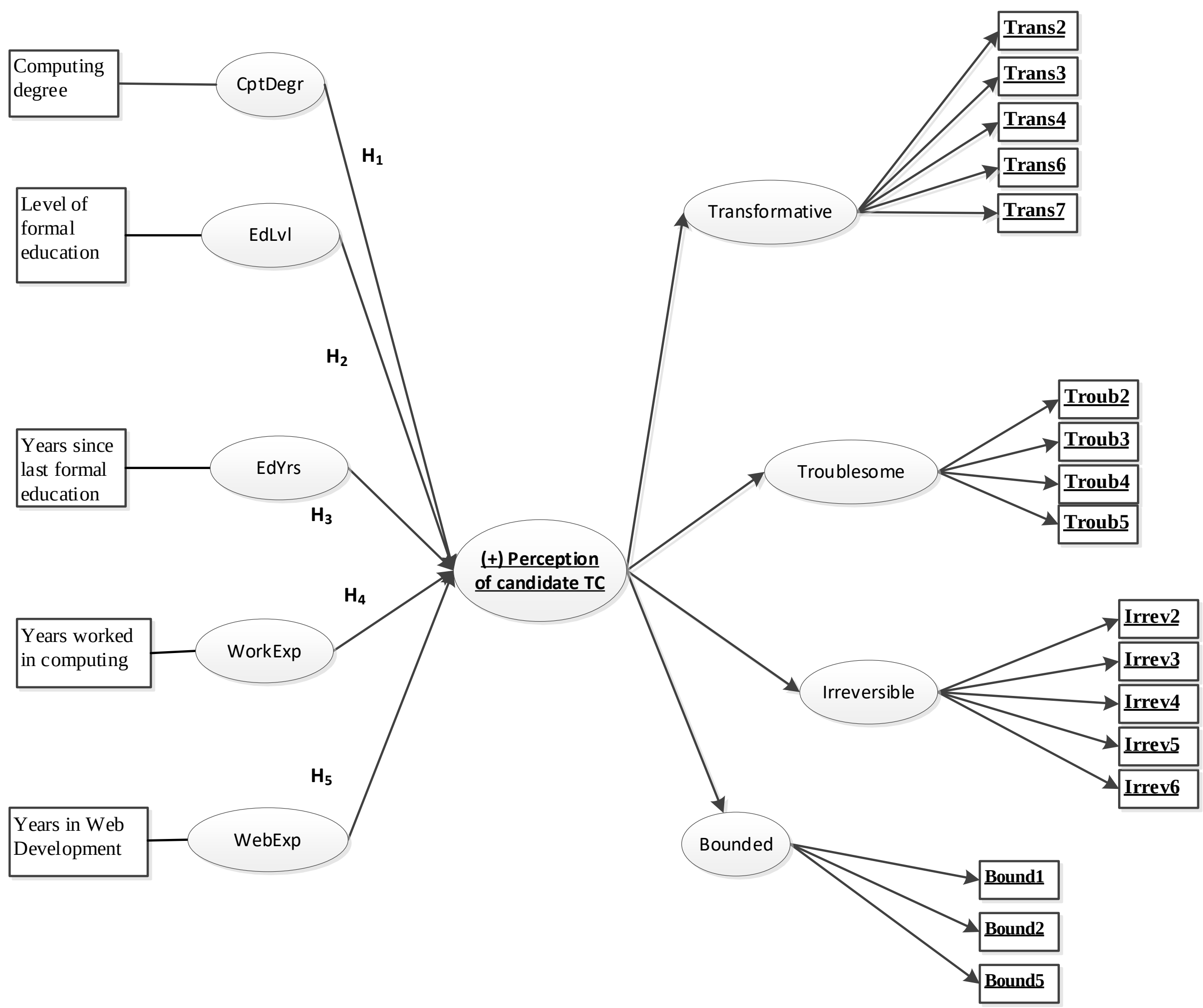

Figure 9. Revised Research Model

PLS-SEM emphasizes explained variance and the establishment of all path estimates (Hair et al., 2017). Therefore, we begin interpretation of the structural model by examining the R^2 values for each endogenous variable. The results indicated that the perception of the concept

SCPB as a TC explained 74.1%, 86.1%, 83.2%, and 78.2% of the variance in all four endogenous variables, Troublesome, Transformative, Irreversible, and Bounded, respectively, and this is considered as a strong effect.

The second item of interest is the significance of the path coefficients, which was tested via examination of their respective t-values. The bootstrapping analysis suggested a significant and positive relationship for all four: between the TC and Troublesome (path = 0.861, t = 40.707, p = 0.000), TC and Transformative (path = 0.928, t = 63.123, p = 0.000), TC and Irreversible (path = 0.913, t = 56.903, p = 0.000), and finally TC and Bounded (path = 0.885, t = 57.345, p = 0.000), thus supporting the theory that a TC is recognized as containing the proposed characteristics of Troublesome, Transformative, Irreversible, and Bounded.

A bootstrapping analysis was completed on measures related to the hypotheses proposed in this research. The analysis produced path coefficients which determined that having a degree in a computing field (path = -0.124, t = 2.014, p = 0.045), the level of education achieved (path = 0.216, t = 2.329, p = 0.020), having work experience in the computing field (path = -0. 294, t = 3.287, p = 0.001), and having Web development experience (path = 0.429, t = 4.719, p = 0.000), all were significant in influencing the perception of Web development professionals, with respect to recognizing the Separation of Content, Presentation, and Behavior (SCPB).

As the five hypotheses were split into paired sub-hypotheses to indicate a positive or negative influence, each pair was expected to be either supported or not supported in alternative directions. Only the Years since Last Formal Education (path = -0.099, t = 0.882, p = 0.376) was not shown to be significant and thus not supportive in either a positive or negative direction; that is, the null hypothesis of having no correlation to the perception of the TC could not be rejected. Age (path = 0.177, t = 2.071, p = 0.045), although not used in a hypothesis, also was shown to be significant. Table 17 displays a summary of all hypotheses, along with their corresponding findings.

Table 17: Hypotheses and Findings Summary

No.	Hypothesis	Finding
H1a	For Web development professionals, having a degree in computing has a positive correlation to their perception of the candidate concept as a TC	Not Supported
H1b	For Web development professionals, having a degree in computing has a negative correlation to their perception of the candidate concept as a TC	Supported
H2a	The level of formal education of Web development professionals has a positive correlation to their perception of the candidate concept as a TC	Supported
H2b	The level of formal education of Web development professionals has a negative correlation to their perception of the candidate concept as a TC	Not Supported
H3a	For Web development professionals, the years since they last were in formal education has a positive correlation to their perception of the candidate concept as a TC	Not Supported
H3b	For Web development professionals, the years since they last were in formal education has a negative correlation to their perception of the candidate concept as a TC	Not Supported
H4a	For Web development professionals, the years they have in computing has a positive correlation to their perception of the candidate concept as a TC	Not Supported
H4b	For Web development professionals, the years they have in computing has a negative correlation to their perception of the candidate concept as a TC	Supported
H5a	For Web development professionals, the years they have in Web development has a positive correlation to their perception of the candidate concept as a TC	Supported
H5b	For Web development professionals, the years they have in Web development has a negative correlation to their perception of the candidate concept as a TC	Not Supported

CHAPTER 5: SUMMARY AND RECOMMENDATIONS

This purpose of this research was to examine the perceptions of professional Web developers of a proposed Threshold Concept, by looking at their agreement with some suggested characteristics of a Threshold Concept, as given by the Threshold Concept theory, and to explore whether their agreement is influenced, either positively or negatively, by experience and education. A literature search determined that all previous research principally utilized qualitative approaches and that no prior quantitative data existed for use, even with potential modification. A predicate requirement then was to develop and validate a scale for determining such perceptions. Such a scale was developed by reviewing the literature for appropriate characteristics and candidate concepts, and by utilizing professionals to review and pre-validate the scale. A national randomized survey was done of professionals working in Web development with the results used in an SPSS analysis to determine factors and reduce the scale items. A second randomized national survey of the same demographic was used with a confirmatory analysis using PLS-SEM.

This final chapter presents a discussion of the findings, limitations, and recommendations for future research.

5.1 Key Findings and their Importance

This research contributes to the literature of the Scholarship of Teaching and Learning by the development of a quantitative approach for the testing of candidate concepts and their concomitant characteristics, as presented by the theory of Threshold Concepts. As previous research has been all qualitative, having a quantitative path for exploration enhances the richness of the research while supporting the theory in the face of certain critiques which purported a lack of rigor in research on this topic.

While qualitative research is useful for determining the characteristics of a TC, this research provides a useful survey tool which potentially can be used with other candidate TCs. The scale should need only minimal modifications, as it was developed to be oriented towards the characteristics rather than the specific candidate TC.

This research also contributed by providing evidence that, at a minimum, the four characteristics examined (Troublesome, Transformative, Irreversible, and Bounded) do influence a concept being a TC.

Further, this study makes a significant contribution by indicating that the candidate TC of Separation of Content, Presentation, and Behavior (SCPB) is considered as a TC by professional Web developers as it sufficiently met the criteria of being Troublesome, Transformative, Irreversible, and Bounded.

With respect to the hypotheses of certain demographic factors influencing perception, the goal was to determine if such factors had a positive or a negative correlation to the perception of the concept as a TC. That goal was accomplished, finding significance in all but one instance, that of the length of time since participants were last in formal education. The null hypothesis in that instance cannot be rejected, so we cannot present any further analysis on this item. As to what the direction, either positive or negative, may mean for each of the other demographics, this was not a research goal. We can only present possibilities based upon the data available and suggest potential directions for future research.

All educational levels were inclined towards perceiving SCPB as a TC. The significant difference between those with graduate degrees and those at a lower level may be indicative of graduates having a better grasp of the bounded, transformative, and irreversible characteristics of

SCPB, as well as being troubling when they learned or even when others are learning, particularly if they may have tried teaching in this area.

The results of work experience in computing and those of experience in Web development were both very significant, however that they trended in the opposite direction was unexpected. Computing work experience also was much weaker in effect than Web development, which was the strongest of all the demographics. This could be due to Web development naturally being separated into job functions (e.g., front-stack, back-stack, etc.) along the lines of SCPB and thus providing more familiarity with the concept. Those with less than one year of Web development experience (M=3.261) differed very significantly with all other groups, and those with one to two years of experience (M=3.887) differed very significantly with those having more than ten years (M=4.265). This reinforces the thought that as Web development experience increases, participants recognized the concept of SCPB as having TC characteristics. That computing experience overall trended negative suggests a need for more investigation.

While PLS results indicate having a degree in a computing field to have a negative relationship to recognition of SCPB as a TC, recall that PLS has no assumption of normality, as does a t-test, and the bootstrapping method is testing the direction of effect that may significantly affect interpretation of the variable. When PLS results are borderline, they can indicate an opposite effect. Noting that the t-test results show a high variation between those with a computing degree (M=3.9, SD=0.72) and those without (M=4.033, SD=0.805), confidence intervals are large, so although the sampling means are slightly different, we cannot rule out that the population means are the same. Consequently, this will require more investigation to determine the impact of having versus not having a degree in computing on perception of SCPB as a TC. Additionally, the numbers of practitioners lacking a degree in a computing field may change over time, and perhaps is an indicator of the relative newness of Web development when compared to other computing areas. However, this raises the question from a teaching perspective of how to ensure that non-computing majors who work in this field may absorb needed concepts.

Although not one of the hypotheses, age showed a very significant difference for those under 30 (M=3.752, SD=0.612) versus those in their 40's (M=4.168, SD=0.68) and a significant difference for those in their 40's versus those 50 and older (M=3.830, SD=1.126). Those in the 30's were not significant (M=3.941, SD=0.782). It is worth noting that the means increased until the 50+ group, but this group had a high variance. Additional future analysis of the differences in responses to the individual characteristics may be worthwhile in determining if longevity in age and perhaps experience affected perception of them versus just the overall result of a TC.

5.2 Theoretical and Practical Implications

The outcomes of this study provide some insights for the TC theory and for continued research in this area. There also is impact for those working in the field of Web development, particularly for those in educational positions, whether within industry or in higher education. Practitioners themselves may benefit, especially as they seek to hire and incorporate the next generation into professional practice.

5.2.1 Implications for Theory

The contribution of this study to research in the Information Technology field within the Scholarship of Teaching and Learning is important in several aspects. First, it quantitatively verified that SCPB is identified as a TC by those who work in Web development. Doing so establishes that the candidate TC of SCPB can now be considered as an actual TC, not just a candidate. By verifying the theory utilizing a quantitative methodology, it also helps to argue against the critics who have complained of lack of rigor in the qualitative approach. Additionally, this study adds the important perspective of practitioners to those of educators and students, who have been the principal focus of previous research. This aspect alone may help to stimulate inclusion of this underutilized group within the bounds of future studies. This research also identified four of the proposed characteristics of a TC as being present and influential for contributing to a concept being a TC, specifically the Troublesome, Transformative, Irreversible, and Bounded attributes. It did not address other proposed characteristics, and these remain open for future study. Of further use is that the current instrument provides a base from which future studies of additional candidate TCs can be developed. This instrument also may be of use in the development of future studies with respect to additional proposed characteristics of a TC beyond the four addressed here. Finally, this study gave an indication of the importance of experience and education in the recognition of a TC, while pointing toward areas of additional research.

5.2.2 Implications for Practice

A practical implication of this research to the IS field is to note that, being an identified TC, SCPB is a concept of which practitioners of Web development will need to have a firm understanding if they are to exhibit professional ways of thinking and practicing. There is an immediate application to instructional staff regarding the importance of explicitly confirming that this TC is assimilated by students. Such concern applies equally for industrial trainers and for those in higher education. This further implies that those in industry who would employ new practitioners may need to verify that potential new hires understand this concept of SCPB. If not, then new hires will have to have an opportunity to learn quickly while on the job. The expectation would be that recruiters would look to higher education to provide new entries into the work force who are already endued with an understanding of such critical concepts. This, in turn, would raise expectations that a degree in higher education, and likely in a relevant field, will continue to increase as a prerequisite to such employment.

5.3 Limitations and Future Research Opportunities

Common method bias is cited as a concern with self-reported scales in survey research (N. R. Sanders, 2008). This can happen when the construct measures are affected due to the structure of the survey instrument, as in its order or grouped structure providing for respondents to correlate answers to questions. To minimize, this survey used a purposeful intermixing in the ordering of the measures for each group. However, this bias may be unavoidable in a survey where self-reported scales are used. Future research may avoid by using a multiple methods approach by combining qualitative and quantitative methods, such as surveys with interviews, thus enhancing reliability of the findings.

Of major concern was the significant length of the survey, despite attempts to limit the number of measures. Longer surveys can result in later survey items receiving quicker and more uniform responses, or no answers at all, as participants increase speed to complete or quit answering altogether (Dillman et al., 2014; Galesic & Bosnjak, 2009). Future studies could try a shorter version, perhaps with no more than five selected measures, limiting the total to no more than twenty.

This study did not include several other candidate TCs, results of which could be evaluated in comparison to those of this TC. Similar results would lend credence to the reliability of the survey. Additional quantitative work is needed, both in using these selected measures and in developing new measures, especially for characteristics not addressed here, in addition to applying both to other candidate TCs.

Although the majority of practitioners have at least an undergraduate degree, there still is a considerable percentage of practitioners who do not. That may be changing, but it can be an area of inquiry for future research.

This study was restricted to the US. The study also was limited to aggregated respondents to obtain a cross-section of Web developers nationwide. An opportunity may exist to try using various firms doing Web development, either internally or for hire, to see if there are any industry effects. Additionally, specific countries and/or regions could be utilized to see if any differences are localized.

5.4 Conclusion

TC is a relatively new framework for identifying concepts that are critical to learners incorporating the "ways of thinking and practice" (WTP) in a subject area. Such TCs are more than just discipline-specific core concepts, in that they entail seeing things in a new way and have a higher requirement for learners to internalize. These grew out of a research project to develop conceptual frameworks to guide in the development of teaching-learning environments, and to evaluate certain concepts in high quality outcomes of learning. While a number of concepts have been identified as potential TCs, all previous research has been pursued utilizing qualitative methods, for the purpose of identifying such candidate TCs. Although several characteristics have been postulated as present, or even potentially required, in a TC, the determination of whether a given candidate TC is, by definition, a TC had yet to be done. The lack of a quantitative approach to verify such candidates has also attracted some criticism. Research to determine candidate TCs has been done using students and faculty, however the inclusion of practicing professionals in the field of Web development is conspicuously absent. In view of TCs being critical concepts for learners to integrate WTP pursuant to becoming professionals in the field, obtaining the perceptions of those in current practice appears to be a needed and missing perspective. Given that previous research utilized the experience of teaching faculty and the educational exposure of students, we postulated that working professionals enjoy aspects of both. We also looked at various candidate TCs to determine which, if any, might best apply to Web development. Therefore, we identified as a research issue the determination of whether a certain candidate TC is perceived as such by Web development professionals and the effects of educational and experiential background on their perception. To accomplish this goal, we established these research questions:

1) To what degree is the candidate TC perceived as such by Web development professionals?
2) How are the perceptions of the TC influenced by the educational backgrounds of Web development professionals?
3) How are the perceptions of the TC influenced by the experiential backgrounds of Web development professionals?

To address the first research question, we selected several of the initially identified characteristics of a TC which have continued to appear in the literature. While we recognize that additional potential characteristics have been theorized, we chose those with an established history in the research. A literature review was unable to identify either an existing scale to address this question nor any which might, with some modification, contribute to our scale development. Therefore, the construction of a survey instrument had to begin with the creating of individual measures. These measures were, at least in part, fashioned by looking at the previous qualitative research and deriving statements reflecting their results. Additionally, these statements were put through an initial refining by use of the Q-sort method until an acceptable set was achieved. An exploratory factor analysis using SPSS identified loadings on four constructs. This allowed for results from a second survey to be used in a confirmatory factor analysis using PLS-SEM for model development. Various statistical tests were applied to ascertain statistical power, reliability, validity, and model measurement. The results illustrate that professional Web developers did perceive the selected candidate TC of SCPB as being a TC by positive and significant responses to the measures for latent constructs Troublesome, Transformative, Irreversible, and Bounded.

For the second research question, certain demographic questions were analyzed from the second survey respondents regarding their level of education completed, whether their education included a degree in the computing field, and the length of time since they were last in formal education. The last item was the only one which did not indicate significance. Education level was particularly significant between those with a graduate degree and those with only undergraduate work or less. While PLS indicated significance and a slightly negative relationship between perception of SCPB as a TC and having a computing degree, a t-test indicated no significant difference between those with and without a computing degree. Given the high variances revealed in the t-test and the slightly negative relationship from PLS, it was noted that additional investigation is needed in this area to determine the impact of having a computing degree.

Finally, the third research question was examined by the additional demographic questions of years of computing experience and years specifically of Web development experience. The responses indicated a very significant correlation for both, but negative for work experience and a much stronger and positive correlation for Web development experience. The negative trend for computing experience in general indicated a need for further investigation.

The strong positive for Web development indicated a potential maturing view of the impact of the TC and possibly the greater recognition of the SCPB in Web development than for computing in general due to job divisions. This also could provide an avenue for future investigation. As a side observation, although not included in the hypotheses, age was found to have a positive and significant correlation.

Overall, this study illustrates that the framework of TCs can be undergirded by quantitative studies with results which confirm the nature of TCs as defined by the theory. In this case, the characteristics of Troublesome, Transformative, Irreversible, and Bounded were shown to be recognized as components of a TC, specifically the SCPB, by the target audience of professional Web developers. It expanded a little-used target population for evaluation of TCs, that being working professionals in the subject field. Further, this dissertation provides a validated survey tool as researchers seek to develop quantitative instruments for investigation of other candidate TCs. Continued quantitative work will add significance to the previous qualitative work, as it will address the critiques calling for additional rigor. Based on the findings of this study, researchers can expect to be able to quantitatively confirm the significance of the presented characteristics in other candidate TCs. Additionally, findings of this dissertation can help researchers and practitioners as they seek to incorporate quantitative evaluations of other proposed characteristics of TCs and, potentially, the effect of the demographics utilized herein.

APPENDICES

APPENDIX A: Survey Instrument

Threshold Concepts in Web Development Survey

Purpose of Study: to determine the perceptions of working Web software development professionals towards Threshold Concepts and whether their perceptions are modified by education or experience.

Definition of Threshold Concept (TC): a concept representing a transformed (changed) way of understanding, or interpreting, or viewing knowledge which is necessary for a learner to progress in ways of thinking and acting like a professional.

Demographic Information:

Birth year: _____

Gender: __ Male __ Female

Work location Zip code: __________

Current employment status:
Full-time (40 hours or more/week)
Part-time (up to 39 hours/week)
Self-employed
Other: __________

My primary job type is : ___front-stack (front-end, user side) ___back-stack (server side) ___full-stack (both front and back) ___none of these

My primary job function is: _________________________________

How many years work experience do you have in the computing field?

< 1 year 1-2yrs 3-5 yrs 6-10 yrs > 10 yrs

How many years work experience do you have specifically in Web development?

< 1 year 1-2yrs 3-5 yrs 6-10 yrs > 10 yrs

How much of your work in Web development within the last 5 years consists of writing code?

<10% 10%-25% 25%-50% 50%-75% >75%

Please select **all** levels of post-secondary education completed and to the right of each, please enter your **area(s)** (majors, concentrations, etc.) If no post-secondary education applies, please select **None**.

__ None
__ Some College (but no degree or certificate) **please enter area of study/major/minor**: ___________________
__ Certificate; **please enter area of study and/or certificate title**: ___________________
__ Associate's degree: **please enter degree** (e.g., AA, AS, etc.) **<u>and</u> area of study/major/minor**: ___________________
__ Bachelor's degree: **please enter degree** (e.g., BA, BS, etc.) **<u>and</u> area of study/major/minor**: ___________________
__ Master's degree: **please enter degree** (e.g., MA, MS, MEd, etc.) **<u>and</u> area of study/major/minor**: ___________________
__ Professional degree: **please enter degree** (e.g., MD, DDS, JD, etc.) **<u>and</u> area of study/major/minor**: ___________________
__ Doctoral degree: **please enter degree** (e.g., PhD, EdD, etc.) **<u>and</u> area of study/major/minor**: ___________________

Year last in school (for latest degree, or last attended if no degree), 4 digits: _______

A particular concept is presented along with a definition, followed by a set of statements to be read in regard to the concept. The statements may or may not describe your beliefs about learning Web development. Please rate each statement according to the following rating scale:

Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree

If you do not understand a statement, or choose not to answer, select "Skip".
If you understand, but have no opinion, choose "Neutral".

<u>Separation of Content, Presentation and Behavior:</u> content typically supplied interactively via a database, with presentation using one technology (e.g., HTML and CSS) and behavior another (e.g., JavaScript). Often expressed as the Model-View-Controller (MVC) design pattern.

1. I am familiar with the concept of Separation of Content, Presentation, and Behavior (or MVC) even if I know it by another name.

   Strongly Agree Agree Neutral Disagree Strongly Disagree

2. In my introduction to Separation of Content, Presentation, and Behavior, I initially thought it seemed a bit unclear or confusing, even if only for a short time.

   Strongly Agree Agree Neutral Disagree Strongly Disagree

3. I would expect professionals outside of software development to have little or no interest in Separation of Content, Presentation, and Behavior.

   Strongly Agree Agree Neutral Disagree Strongly Disagree

4. Separation of Content, Presentation, and Behavior was important to my integrating the parts of Web development into a "bigger picture".

   Strongly Agree Agree Neutral Disagree Strongly Disagree

5. My use of Separation of Content, Presentation, and Behavior has transformed to be more instinctive than when I first learned about it.

   Strongly Agree Agree Neutral Disagree Strongly Disagree

6. Like riding a bike, my understanding of Separation of Content, Presentation, and Behavior is <u>not</u> likely to be "unlearned" (completely forgotten.)

   Strongly Agree Agree Neutral Disagree Strongly Disagree

7. After I first encountered Separation of Content, Presentation, and Behavior, I thought that I understood it, but then I had some difficulty solving problems using this concept.

   Strongly Agree Agree Neutral Disagree Strongly Disagree

8. I would <u>not expect</u> someone outside of computing to know about Separation of Content, Presentation, and Behavior.

   Strongly Agree Agree Neutral Disagree Strongly Disagree

9. Understanding Separation of Content, Presentation, and Behavior altered my perception of Web development.

   Strongly Agree Agree Neutral Disagree Strongly Disagree

10. I can converse better in other aspects of Web development by understanding Separation of Content, Presentation, and Behavior.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

11. Understanding Separation of Content, Presentation, and Behavior enables me to work more quickly at bringing different development parts together for a task.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

12. When I first learned about Separation of Content, Presentation, and Behavior, the concept did not seem obvious.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

13. My understanding of Separation of Content, Presentation, and Behavior will last for however long I might need to work with it.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

14. My understanding of doing Web development was modified once I understood Separation of Content, Presentation, and Behavior.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

15. Understanding Separation of Content, Presentation, and Behavior allows seeing connections which enable better job performance.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

16. Understanding Separation of Content, Presentation, and Behavior is more important for computing professionals than for non-computing professionals.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

17. I would expect anyone learning about Separation of Content, Presentation, and Behavior to be a bit confused at first.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

18. Learning about Separation of Content, Presentation, and Behavior changed my thinking on how to do Web development.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

19. My understanding of Separation of Content, Presentation, and Behavior will remain and not revert to a previous perception.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

20. Knowledge of Separation of Content, Presentation, and Behavior helps to understand connections [interrelationships] in my discipline that otherwise might not be noticed.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

21. I can integrate other aspects of software development (and possibly other computing areas) more easily by understanding Separation of Content, Presentation, and Behavior.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

22. Understanding Separation of Content, Presentation, and Behavior typically is within the boundaries of software development, as opposed to non-computing areas.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

23. My understanding of Separation of Content, Presentation, and Behavior remains permanent enough to explain it in my own words.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

24. My thinking transformed from more like a learner to more like that of a professional after understanding Separation of Content, Presentation, and Behavior.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

25. While learning about Separation of Content, Presentation, and Behavior, I found it confusing to my original understanding of software development.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

26. After I first thought that I understood the concept of Separation of Content, Presentation, and Behavior, how to apply it (i.e., transfer learning to practice) was harder than expected.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

27. Learning about Separation of Content, Presentation, and Behavior showed to me concepts which are bounded (mostly unique) to my area of work.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

28. Knowing Separation of Content, Presentation, and Behavior helps me to understand how my code integrates into a larger project.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

29. I can recall altering my understanding of how Separation of Content, Presentation, and Behavior relates to the "real world" (i.e., my job).

    Strongly Agree Agree Neutral Disagree Strongly Disagree

30. After I learned about Separation of Content, Presentation, and Behavior, I gained a changed view of Web development.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

31. Upon learning about Separation of Content, Presentation, and Behavior, my ideas about how to program changed significantly.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

32. Having understood Separation of Content, Presentation, and Behavior, I can continue to use it without forgetting how.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

33. I can use Separation of Content, Presentation, and Behavior appropriately because my understanding of it is irreversible.

    Strongly Agree Agree Neutral Disagree Strongly Disagree

Thank you for participating!

APPENDIX B. Sample Item Placement Ratio Calculation

As an example, we illustrate with a simple case having four constructs and fifteen items developed as potential measures for each. Given two judges on a panel, there exists a possible 30 placements per construct, for a total of 120 placements. An actual versus theoretical matrix could be constructed as shown in Table **18**, including an actual “not applicable” column, listed as N/A:

Table 18: Sample Actual versus Theoretical Matrix

CONSTRUCTS		Actual						
		A	B	C	D	N/A	Total	% Hits
Theoretical	A	26	2	1	0	1	30	87%
	B	0	20	10	0	0	30	67%
	C	0	2	27	1	0	30	90%
	D	0	0	0	30	0	30	100%
Item placements, 120; hits, 103 (within the oval); overall hit ratio, 85.8%.								

The item placement ratio is the sum of the items placed by the judges into the intended categories, which is the sum of items in the indicated diagonal as shown in Table **18**Table 18: Sample Actual versus Theoretical Matrix, divided by the total number of possible placements. In this case, the theoretical maximum number of placements is 120 (4 constructs at 15 placements per judge, with 2 judges), with a total of 103 “hits”, achieving an overall hit ratio of 86% (rounded). Even more importantly, one can see that row D indicates that all 30 of the item placements were in the target construct. However, row B shows only 67% are within the target. Moreover, with one-third (10) of the items are placed within C, possibly indicating that the underlying items for B and C are not sufficiently differentiated. This leads to more confidence with a scale based on C, but more hesitancy on a scale based upon B. Examination of the off-diagonal entries leads to an indication of the complexity of a given construct. Actual constructs with a high number in the off-diagonal entries may be considered as too ambiguous.

APPENDIX C. Measurement Statements by Construct in Q-Sorts

Round 1:

Troublesome

1. After I first encountered [TC], I thought that I understood it, but then I had some difficulty solving problems using this concept.
2. In my introduction to [TC], I initially thought it seemed a bit unclear or confusing, even if only for a short time.
3. When I first learned about [TC], the concept did not seem obvious.
4. I would expect anyone learning about [TC] to be a bit confused at first.
5. I had to incorporate a new understanding of software development while learning about [TC].
6. I understood the concept of [TC], but applying it was considerably more difficult.
7. Upon learning about [TC], my ideas about how to program changed significantly. (1)
8. Learning to write succinct, structured, error-free code takes practice.

Bounded

1. Understanding [TC] enables me to perform more like a professional.
2. I would expect someone who is not in software development to know about [TC].
3. Understanding [TC] is more important for computing professionals than for non-computing professionals.

4. I would expect professionals outside of software development to have little or no interest in [TC].
5. Understanding [TC] is a trait of software development professionals.
6. Learning about [TC] helped me to see principles that uniquely apply to my type of work.

Integrative

1. [TC] was a key concept to my understanding of Web development.
2. Understanding [TC] helps me to converse better in other aspects of Web development.
3. Understanding [TC] increases productivity.
4. Understanding [TC] improves my job performance.
5. Knowledge of [TC] helps to understand connections [interrelationships] in my discipline that otherwise might not be noticed.
6. I can "fit" pieces of Web development (and possibly other computing areas) together more easily by understanding [TC].
7. Having learned about [TC], I can better see the impact of my code on a project.

Transformative

1. As I approach a programming application, [TC] comes more naturally to me now than when I first learned about it.
2. Understanding [TC] makes it easier to do Web development.
3. Understanding [TC] enables me to work more quickly. (2)

4. Understanding [TC] now seems straightforward.
5. Learning about [TC] changed how I thought about doing Web development.
6. My understanding of doing Web development changed once I understood [TC].
7. Compared to when I first encountered [TC], it now seems clearer to me.
8. I now understand [TC] better than when I first learned about it.
9. I can recall when it dawned on me how [TC] relates to the "real world" (i.e., my job).
10. I gained new insights into software development when I learned about [TC].
11. I think more like a professional now after understanding [TC].
12. Once I grasped an understanding of the basics of [TC], I was able to confidently use it in order to produce a better work product. (3)

Irreversible

1. Understanding [TC], like riding a bike, once learned it is not easily "unlearned" (completely forgotten.)
2. I will continue to understand [TC] throughout my career.
3. I do not have to refresh my understanding of [TC] before applying it to a programming application.
4. I can explain [TC] in my own words.
5. I am able to use [TC] without much difficulty.
6. I am able to write code that can be reused for other applications. (4)
7. I am able to easily make use of reusable code (i.e., previously created code, as in code libraries.) (4)

(1) Moved after Q-Sort Round 1to Transformative
(2) Moved after Q-Sort Round 1 to Integrative
(3) Moved after Q-Sort Round 1 to Irreversible
(4) Deleted after Q-Sort Round 1

Round 2:

Troublesome

1. After I first encountered [TC], I thought that I understood it, but then I had some difficulty solving problems using this concept.
2. In my introduction to [TC], I initially thought it seemed a bit unclear or confusing, even if only for a short time.
3. When I first learned about [TC], the concept did not seem obvious.
4. I would expect anyone learning about [TC] to be a bit confused at first.
5. While learning about [TC], I found it confusing to my original understanding of software development.
6. After I first thought that I understood the concept of [TC], how to apply it (i.e., transfer learning to practice) was harder than expected.
7. Learning to write succinct, structured, error-free code was troublesome.

Bounded

1. Understanding [TC] tends to identify me as a computing professional.
2. I would not expect someone outside of computing to know about [TC].

3. Understanding [TC] is more important for computing professionals than for non-computing professionals.
4. I would expect professionals outside of software development to have little or no interest in [TC].
5. Understanding [TC] typically is within the boundaries of software development, as opposed to non-computing areas.
6. Learning about [TC] showed to me concepts which are bounded (mostly unique) to my area of work.

Integrative

1. [TC] was important to my integrating the parts of Web development into a "bigger picture".
2. I can converse better in other aspects of Web development by understanding [TC].
3. Understanding [TC] increases productivity.
4. Understanding [TC] enables better job performance.
5. Knowledge of [TC] helps to understand connections [interrelationships] in my discipline that otherwise might not be noticed.
6. I can integrate other aspects of software development (and possibly other computing areas) more easily by understanding [TC].
7. Knowing [TC] helps me to understand how my code integrates into a larger project.
8. Understanding [TC] enables me to work more quickly.

Transformative

1. My use of [TC] has changed to be more instinctive than when I first learned about it.
2. Understanding [TC] altered my perception of Web development.
3. Understanding [TC] now seems straightforward.
4. Learning about [TC] changed how I thought about doing Web development.
5. My understanding of doing Web development changed once I understood [TC].
6. Compared to when I first encountered [TC], it now seems clearer to me.
7. My understanding of [TC] is transformed, compared to before learning of its use in developing software.
8. I can recall changing my understanding of how [TC] relates to the "real world" (i.e., my job). (1)
9. After I learned about [TC], I gained a changed view of Web development.
10. My thinking transformed to be more like that of a professional after understanding [TC].
11. Upon learning about [TC], my ideas about how to program changed significantly.

Irreversible

1. Having understood [TC], I can continue to use it successfully.
2. My understanding of [TC], like riding a bike, it is not likely to be "unlearned" (completely forgotten.)
3. I will continue to understand [TC] throughout my career.
4. I do not have to refresh my understanding of [TC] before using it.

5. My understanding of [TC] is sufficient to explain it in my own words, without reviewing beforehand.
6. I can use [TC] appropriately because my understanding of it is irreversible.

(1) Moved after Q-Sort Round 2 to Integrative

Round 3:

Troublesome

1. After I first encountered [--------], I thought that I understood it, but then I had some difficulty solving problems using this concept.
2. In my introduction to [--------], I initially thought it seemed a bit unclear or confusing, even if only for a short time.
3. When I first learned about [--------], the concept did not seem obvious.
4. I would expect anyone learning about [--------] to be a bit confused at first.
5. While learning about [--------], I found it confusing to my original understanding of software development.
6. After I first thought that I understood the concept of [--------], how to apply it (i.e., transfer learning to practice) was harder than expected.
7. Learning to write succinct, structured, error-free code can be troublesome. (1)

Bounded

1. Understanding [--------] tends to identify me as a computing professional, as compared to non-computing disciplines. (1)
2. I would not expect someone outside of computing to know about [--------].

3. Understanding [--------] is more important for computing professionals than for non-computing professionals.
4. I would expect professionals outside of software development to have little or no interest in [--------].
5. Understanding [--------] typically is within the boundaries of software development, as opposed to non-computing areas.
6. Learning about [--------] showed to me concepts which are bounded (mostly unique) to my area of work.

Integrative

1. [--------] was important to my integrating the parts of Web development into a "bigger picture".
2. I can converse better in other aspects of Web development by understanding [--------].
3. Understanding [--------] increases productivity, as I can better integrate all of the technology "pieces" into a whole. (1)
4. Understanding [--------] allows seeing connections which enable better job performance.
5. Knowledge of [--------] helps to understand connections [interrelationships] in my discipline that otherwise might not be noticed.
6. I can integrate other aspects of software development (and possibly other computing areas) more easily by understanding [--------].
7. Knowing [--------] helps me to understand how my code integrates into a larger project.
8. Understanding [--------] enables me to work more quickly at bringing different development parts together for a task.

9. I can recall altering my understanding of how [--------] relates to the "real world" (i.e., my job).

Transformative

1. My use of [--------] has transformed to be more instinctive than when I first learned about it.
2. Understanding [--------] altered my perception of Web development.
3. Understanding [--------] now seems more informed, compared to my previous perception. (2)
4. Learning about [--------] changed my thinking on how to do Web development.
5. My understanding of doing Web development was modified once I understood [--------].
6. Compared to when I first encountered [--------], my understanding has changed. (1)
7. My understanding of [--------] is different, compared to first learning of its use in software development. (1)
8. After I learned about [--------], I gained a changed view of Web development.
9. My thinking transformed from more like a learner to more like that of a professional after understanding [--------].
10. Upon learning about [--------], my ideas about how to program changed significantly.

Irreversible

1. Having understood [--------], I can continue to use it without forgetting how.
2. Like riding a bike, my understanding of [--------] is not likely to be "unlearned" (completely forgotten.)

3. My understanding of [--------] will last for however long I might need to work with it.

4. My understanding of [--------] will remain and not revert to a previous perception.

5. My understanding of [--------] remains permanent enough to explain it in my own words.

6. I can use [--------] appropriately because my understanding of it is irreversible.

(1) Deleted after Q-Sort Round 3 and before Final Survey

APPENDIX D. Cohen's Kappa

Cohen's Kappa (Hair et al., 2017) is a coefficient measuring the degree of agreement in nominal scales. This statistic thus is a measure of interrater reliability, the degree to which judges give the same scoring to the same data item. This statistic assumes that two raters, or judges, are specifically chosen. If more than two are chosen from a rater population, Fleiss' kappa should be used instead (Cohen, 1960). The Kappa statistic is the proportion of agreement after removing chance agreement as a consideration.

Cohen's Kappa typically varies from 0 to 1. While not likely, a negative value would indicate less than chance agreement and be of no further interest. Although there is no generally required score, (Fleiss, 1971) provides a table, reproduced below in Table **19**, as a useful "benchmark", although admittedly with arbitrary divisions.

Table 19: Kappa Values

Kappa Statistic	Strength of Agreement
< 0.00	Poor
0.00 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	Almost Perfect

More recent studies consider levels above 0.65 as acceptable (Landis & Koch, 1977). A brief example of how to calculate the value follows.

As illustrated in Table 20Table **20**: Cohen's Kappa - Sample Table of Frequencies, given two judges who independently rate a set of N items into c categories, let X_{ij} indicate the number in row i and column j. Additional columns and rows can indicate other possible categories.

Table 20: Cohen's Kappa - Sample Table of Frequencies

Raters			<u>Judge 1</u>		
		Categories	c_1	c_2	Totals
Judge 2		c_1	X_{11}	X_{12}	X_{1+}
		c_2	X_{21}	X_{22}	X_{2+}
		Totals	X_{+1}	X_{+2}	N

The above table of frequencies instead can be created using percentages, P, by dividing each numerical entry (observed frequency) by N, resulting in a new table of percentages as shown in Table 21:

Table 21: Cohen's Kappa - Sample Table of Percentages

Raters			<u>Judge 1</u>		
		Categories	c_1	c_2	Totals
Judge 2		c_1	P_{11}	P_{12}	P_{1+}
		c_2	P_{21}	P_{22}	P_{2+}
		Totals	P_{+1}	P_{+2}	100

where Pij is the percentage of components in the ith row and jth column.

This table of percentages is used to describe Cohen's Kappa. From this table, the observed proportional agreement of the judges is calculated as:

$$p_o = \sum_{i=1}^{c} p_{ii}$$

Whereas the expected agreement of the judges by chance is calculated as:

$$p_e = \sum_{i=1}^{c} p_{i+}\, p_{+i}$$

Cohen's Kappa is then calculated as:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

APPENDIX E. Perreault and Leigh's Index of Reliability

Perreault and Leigh's Index of Reliability (I_r) is a measure of interrater reliability in common use which, similarly to Cohen's κ, also considers the observed proportional agreement, p_o (aka item placement ratio, hit ratio), between pairs of judges, and also accounting for the number of construct categories (c). The index, I_r, can be calculated as

$$l_r = \sqrt{\frac{p_o - \frac{1}{c}}{\frac{c}{c-1}}}$$

Where $p_o >= 1/c$. If $p_o < 1/c$, l_r is set to 0.

APPENDIX F. Q-Sort Analysis Results

Table 22: First Round Q-Sort Results and Item Placement Ratio

CONSTRUCTS		Actual Judge Placements							
		Troublesome	Bounded	Integrative	Transformative	Irreversible	N/A	Total	% Hits
THEORECTICAL	Troublesome	12	1	0	2	0	1	16	75%
	Bounded	0	7	3	1	0	1	12	58%
	Integrative	0	0	10	3	1	0	14	71%
	Transformative	0	2	7	11	4	0	24	46%
	Irreversible	1	0	2	3	6	2	14	43%

Item Placements: 80 Hits: 46 Overall Hit ratio: 58%

Table 23: First Round Inter-Judge Raw Agreement (%) for κ and l_r Statistics

CONSTRUCTS		Judge 1						
		Troublesome	Bounded	Integrative	Transformative	Irreversible	N/A	Total
Judge 2	Troublesome	0.125	0	0	0	0	0	0.125
	Bounded	0.025	0.05	0.075	0	0.025	0	0.175
	Integrative	0	0	0.1	0.05	0.1	0	0.25
	Transformative	0	0	0.125	0.125	0.025	0	0.275
	Irreversible	0.025	0	0	0.025	0.025	0	0.075
	N/A	0.025	0.025	0	0.025	0.025	0	0.1
	Total	0.2	0.075	0.3	0.225	0.2	0	1

$p_o = 0.425$ $p_e = 0.19$ $\kappa = 0.290$ $l_r = 0.5567764$

Table 24: Second Round Q-Sort Results and Item Placement Ratio

CONSTRUCTS		Actual Judge Placements							
		Troublesome	Bounded	Integrative	Transformative	Irreversible	N/A	Total	% Hits
THEORECTICAL	Troublesome	13	0	0	0	0	1	14	93%
	Bounded	0	10	2	0	0	0	12	83%
	Integrative	0	0	14	1	2	1	18	78%
	Transformative	0	0	8	10	2	0	20	50%
	Irreversible	1	0	6	0	5	0	12	42%

Item Placements: 76 Hits: 52 Overall Hit ratio: 68%

Table 25: Second Round Inter-Judge Raw Agreement (%) for κ and l_r Statistics

CONSTRUCTS		Judge 1						
		Troublesome	Bounded	Integrative	Transformative	Irreversible	N/A	Total
Judge 2	Troublesome	0.157894737	0	0	0	0	0.026	0.184
	Bounded	0	0.105263	0	0	0	0	0.105
	Integrative	0.026315789	0.052632	0.1842105	0	0.15789474	0.026	0.447
	Transformative	0	0	0.1578947	0.052631579	0.02631579	0	0.237
	Irreversible	0	0	0	0	0.02631579	0	0.026
	N/A	0	0	0	0	0	0	0
	Total	0.184210526	0.1578947	0.34210526	0.052631579	0.210526316	0.0526	1

$p_o = 0.526$ $p_e = 0.22$ $\kappa = 0.391$ $l_r = 0.6569467$

Table 26: Third Round Q-Sort Results and Item Placement Ratio

CONSTRUCTS		Actual Judge Placements							
		Troublesome	Bounded	Integrative	Transformative	Irreversible	N/A	Total	% Hits
THEORECTICAL	Troublesome	13	0	0	0	0	1	14	93%
	Bounded	0	10	0	0	0	2	12	83%
	Integrative	0	0	17	0	1	0	18	94%
	Transformative	1	0	0	17	0	2	20	85%
	Irreversible	0	0	0	0	12	0	12	100%

Item Placements: 76 Hits: 69 Overall Hit ratio: 91%

Table 27: Third Round Inter-Judge Raw Agreement (%) for κ and l_r Statistics

CONSTRUCTS		Judge 1						
		Troublesome	Bounded	Integrative	Transformative	Irreversible	N/A	Total
Judge 2	Troublesome	0.157894737	0	0	0	0	0.026	0.184
	Bounded	0	0.1052632	0	0	0	0	0.105
	Integrative	0	0	0.21052632	0	0.02631579	0	0.237
	Transformative	0.026315789	0	0	0.184210526	0	0	0.211
	Irreversible	0	0	0	0	0.15789474	0	0.158
	N/A	0	0.0526316	0	0.052631579	0	0	0.105
	Total	0.184210526	0.1578947	0.21052632	0.236842105	0.18421053	0.026	1

$p_o = 0.816$ $p_e = 0.18$ $\kappa = 0.775$ $l_r = 0.88258$

APPENDIX G. Nonresponse Bias Analysis Results

Examining for nonresponse bias is important to determine whether or not respondents did indeed participate independently or if in a systematic pattern. To assess the pattern of participation, demographic data collected from the respondents are compared. Data of the respondents from the first collection (early) and from the second collection (late) were subjected to the chi-square (χ^2) test due to the categorical nature of the variables. IBM SPSS 26 was used to perform the analysis. The following tables in this appendix present the results of the χ^2 tests.

Table 28: Gender - Nonresponse Bias Test

Gender	Early Responder	Late Responder	χ^2 test
Male	155	177	$\chi^2 = 2.000$ df = 1 p = 0.157
Female	82	71	

Table 29: Employment Status - Nonresponse Bias Test

Employment Status	Early Responder	Late Responder	χ^2 test
Full-time	203	218	$\chi^2 = 4.565$ df = 3 p = 0.207
Part-time	11	13	
Self-employed	19	17	
Other	4*	0	

* “Other” as a choice should have few, if any, as all were expected to be currently employed.

Table 30: Job Type - Nonresponse Bias Test

Job Type	Early Responder	Late Responder	χ^2 test
Front-stack	68	63	$\chi^2 = 2.599$ df = 3 p = 0.458
Back-stack	19	30	
Full-stack	108	109	
None of these	42	46	

Table 31: Years in Computing Field - Nonresponse Bias Test

Computing Years	Early Responder	Late Responder	χ^2 test
<1	9	11	$\chi^2 = 2.540$ df = 4 p = 0.637
1-2	27	26	
3-5	57	69	
6-10	74	63	
>10	70	79	

Table 32: Years in Web Development - Nonresponse Bias Test

Years in Web	Early Responder	Late Responder	χ^2 test
<1	24	30	$\chi2 = 3.514$ df = 4 p = 0.476
1-2	36	37	
3-5	68	85	
6-10	67	55	
>10	42	41	

Table 33: Percentage of Job Coding - Nonresponse Bias Test

% Job Coding	Early Responder	Late Responder	χ^2 test
<10%	38	33	$\chi^2 = 3.113$ df = 4 p = 0.539
10-25	31	30	
25-50	59	75	
50-75	67	60	
>75%	42	50	

Table 34: Age - Nonresponse Bias Test

Age	Early Responder	Late Responder	χ^2 test
<30	40	48	$\chi^2 = 6.712$ df = 3 p = 0.082
<40	101	78	
<50	77	95	
50+	19	27	

Table 35: Computing Degree - Nonresponse Bias Test

Computing Degree	Early Responder	Late Responder	χ^2 test
Yes	103	99	$\chi^2 = .625$ df = 1 p = 0.429
No	134	149	

Table 36: Post-Secondary Education Level - Nonresponse Bias Test

Education Level	Early Responder	Late Responder	χ^2 test
None	33	37	$\chi^2 = 11.363$ df = 6 p = 0.078
Some college	26	23	
Certificate	6	12	
Associate	16	18	
Bachelor	80	57	
Master	58	85	
Doctoral	18	16	

Table 37: Years Since Last Education - Nonresponse Bias Test

Years Last Educ	Early Responder	Late Responder	χ^2 test
<10	127	137	$\chi^2 = 0.134$ df = 1 p = 0.714
10+	110	111	

APPENDIX H. Composite Reliability

Composite Reliability is calculated by

$$\rho_c = \frac{\left(\sum_{i=1}^{M} l_i\right)^2}{\left(\sum_{i=1}^{M} l_i\right)^2 + \sum_{i=1}^{M} \mathrm{var}\left(e_i\right)}$$

where l_i denotes the loading of a LOC *i* of a specific HOC measured with *M* LOCs ($i = 1, ..., M$), e_i is the measurement error of LOC *i*, and var(e_i) signifies the variance of the measurement error, which by definition is $1-l_i^2$.

APPENDIX I. Cronbach's (standardized) Alpha

The formula for the (standardized) Cronbach's alpha of the HCM is:

$$\text{Cronbach's } \alpha = \frac{M \bullet \bar{r}}{(1 + (M - 1) \bullet \bar{r})}$$

where $\bar{r}$ is the average correlation of the LOCs.

REFERENCES

Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-Centered Design. In *Encyclopedia of Human-Computer Interaction*. Thousand Oaks: Sage Publications.

ACM. (2015). The 2012 ACM Computing Classification System toc. *The 2012 ACM Computing Classification System.* Retrieved from https://www.acm.org/about/class/2012?pageIndex=0

Afthanorhan, W. M. A. B. A. (2013). A Comparison Of Partial Least Square Structural Equation Modeling (PLS-SEM) and Covariance Based Structural Equation Modeling (CB-SEM) for Confirmatory Factor Analysis. *International Journal of Engineering Science and Innovative Technology (IJESIT), 2*(5), 198-205.

Aghaei, S., Nematbakhsh, M. A., & Farsani, H. K. (2012). Evolution of the World Wide Web: from Web 1.0 to Web 4.0. *International Journal of Web & Semantic Technology, 3*(1), 1-10.

Alston, P., Walsh, D., & Westhead, G. (2015). Uncovering "Threshold Concepts" in Web Development: An Instructor Perspective. *ACM Transactions on Computing Education (TOCE) - Special Issue on Web Development, 15*(1), 2:1-2:18.

Ankolekar, A., Krötzsch, M., Tran, T., & Vrandečić, D. (2007). *The two cultures: Mashing up Web 2.0 and the Semantic Web.* Paper presented at the Proceedings of the 16th international conference on World Wide Web, Banff, Alberta, Canada.

Armstrong, J. S., & Overton, T. S. (1977). Estimating Nonresponse Bias in Mail Surveys. *Journal of Marketing Research, 14*(1), 396-402.

Baillie, C., Bowden, J. A., & Meyer, J. H. F. (2013). Threshold capabilities: threshold concepts and knowledge capability linked through variation theory. *High Educ, 65*, 227-246.

Barradell, S. (2013). The identification of threshold concepts: a review of theoretical complexities and methodological challenges. *High Educ, 65*, 265-276.

Baskerville, R., Ramesh, B., Levine, L., Pries Heje, J., & Slaughter, S. (2003). Is internet-speed software development different? *IEEE Software, 20*(6), 70-77.

Bennedsen, J., & Caspersen, M. E. (2007). Failure rates in Introductory Programming. *ACM SIGCSE Bulletin, 39*(2), 32-36.

Bosse, Y., & Gerosa, M. A. (2016). Why is programming so difficult to learn? Patterns of Difficulties Related to Programming Learning Mid-Stage. *ACM SIGSOFT Software Engineering Notes, 41*(6), 1-6.

Boustedt, J., Eckerdal, A., McCartney, R., Moström, J. E., Ratcliffe, M., Sanders, K., & Zander, C. (2007). *Threshold Concepts in Computer Science: Do they exist and are they useful?* Paper presented at the SIGCSE '07 The 38th ACM Technical Symposium on Computer Science Education, Covington, Kentucky, USA.

Boyer, E. L. (1990). *Scholarship Reconsidered: Priorities of the Professoriate*. Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York, NY: Routledge.

Cardellini, V., Colajanni, M., & Yu, P. S. (1999). Dynamic Load Balancing on Web-server Systems. *IEEE Internet Computing, 3*(3), 28-39.

Carey, S. (1991). Knowledge Acquisition: Enrichment or Conceptual Change? In *The Epigenesis of Mind: Essays on Biology and Cognition* (pp. 257-292). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Chen, C., Chen, Y., Miao, H., & Wang, H. (2013). *Usage-pattern based statistical web testing and reliability measurement.* Paper presented at the The 4th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2013) and the 3rd International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH).

Chin, W. W. (1998). Issues and opinion on structural equation modeling. *MIS Quarterly, 22*(1), vii-xvi.

Chin, W. W. (1998). The Partial Least Squares Approach to Structural Equation Modeling. In *Modern methods for business research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37-46.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Routledge.

Connolly, R. (2019). *Facing Backwards While Stumbling Forwards: The Future of Teaching Web Development.* Paper presented at the SIGCSE '19 the 50th ACM Technical Symposium on Computer Science Education, Minneapolis, MN.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*, *10*, 1-9. doi:https://doi.org/10.7275/jyj1-4868

Cousin, G. (2006). An introduction to threshold concepts. *Planet*(17), 4-5.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Davies, P. (2003). *Threshold Concepts: how can we recognise them?* Paper presented at the 10th EARLI Biennial Conference, Padova, Italy.

Davies, P. (2006). Threshold concepts: how shall we recognise them? In J. H. F. Meyer & R. Land (Eds.). London: Routledge.

Davies, P., & Mangan, J. (2005). Recognising Threshold Concepts: an exploration of different approaches. In *European Association in Learning and Instruction (EARLI)*. Nicosia, Cyprus.

Davies, P., & Mangan, J. (2007). Threshold concepts and the integration of understanding in economics. *Studies in Higher Education, 32*(6), 711-726.

Davies, P., & Mangan, J. (2011). Threshold Concepts in Economics: implications for teaching, learning and assessment. In *The Handbook for Economics Lecturers* (pp. 1-29). Staffordshire, UK: The Business School, Staffordshire University.

Deutskens, E., De Ruyter, K., & Wetzels, M. (2006). An assessment of equivalence between online and mail surveys in service research. *Journal of Service Research, 8*(4), 346-355.

Dias, D. M., Kish, W., Mukherjee, R., & Tewari, R. (1996). *A Scalable and Highly Available Web Server.* Paper presented at the 41st IEEE Computer Society Int'l. Conf. (COMPCON'96), Santa Clara, CA.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: the tailored design method*. Hoboken, NJ: John Wiley & Sons.

Dorn, B., & Guzdial, M. (2010). *Discovering Computing: Perspectives of Web Designers.* Paper presented at the Proceedings of the 6th International Workshop on Computing Education Research (ICER’10), Aarhus, Denmark.

Eckerdal, A., McCartney, R., Moström, J. E., Ratcliffe, M., Sanders, K., & Zander, C. (2006). *Putting Threshold Concepts into Context in Computer Science Education.* Paper presented at the ITICSE '06: Proceedings of the 11th annual SIGCSE Conference on Innovation and Technology in Computer Science Education, Bologna, Italy.

Eckerdal, A., McCartney, R., Moström, J. E., Sanders, K., Thomas, L., & Zander, C. (2007). *From Limen to Lumen: Computing students in liminal spaces.* Paper presented at the ICER '07, Atlanta, GA.

Economic and Social Research Council. (2005, 6 30). *Enhancing Teaching-Learning Environments in Undergraduate Courses*. Retrieved from http://www.esrc.ac.uk/my-esrc/grants/L139251099/read

ETL Project. (2005, 6). *Enhancing Teaching-Learning Environments in Undergraduate Courses*. Retrieved from http://www.etl.tla.ed.ac.uk/

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.

Flanagan, M. (Producer). (2015, 6 14). Threshold Concepts: A short introduction and bibliography. *Threshold Concepts: Undergraduate Teaching, Postgraduate Training and Professional Development*. Retrieved from http://www.ee.ucl.ac.uk/~mflanaga/thresholds.html#ML2003

Fleiss, J. L. (1971). Measuring Nominal Scale Agreement Among Many Raters. *Psychological Bulletin, 76*(5), 378-382.

Fuchs, C., Hofkirchner, W., Schafranek, M., Raffl, C., Sandoval, M., & Bichler, R. (2010). Theoretical Foundations of the Web: Cognition, Communication, and Co-Operation. Towards an Understanding of Web 1.0, 2.0, 3.0. *Future Internet, 2*(1), 41-59.

Galesic, M., & Bosnjak, M. (2009). Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly, 73*(2), 349-360. doi:10.1093/poq/nfp031

Geisser, S. (1974). A predictive approach to the random effect model. *Biometrika, 61*(1), 101-107.

Gibson, C. (2006). *Website Engineering in the Real World.* Paper presented at the ACM-SE 44 Proceedings of the 44th annual Southeast regional conference, Melbourne, FL.

Gray, J. (2001). *Functionality, Availability, Agility, Manageability, Scalability -- the New Priorities of Application Design.* Paper presented at the Proceedings HTPS, Asilomar.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis*. Upper Saddle River, NJ: Prentice Hall.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Los Angeles: Sage.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM - Indeed a Silver Bullet. *Journal of Marketing Theory and Practice, 19*(2), 139-151.

Hair, J. F., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2018). *Advanced Issues in Partial Least Squares Structural Equation Modeling*. Los Angeles, CA, USA: Sage Publications, Inc.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(4), 115-135.

Hoar, R., & Connolly, R. (2016). *The Garden of Earthly Delights: Constructing and Revising a Web Development Textbook.* Paper presented at the Proceedings of the 25th International Conference Companion on World Wide Web (WWW '16 Companion), Montréal, Québec, Canada.

Huber, M. (2019). What is the Scholarship of Teaching and Learning? *Teaching Commons - Teaching Talk.* Retrieved from https://teachingcommons.stanford.edu/teaching-talk/what-scholarship-teaching-and-learning-mary-huber

Hulland, J. (1999). Use of Partial Least Squares (PLS) in Strategic Management Research: A Review of Four Recent Studies. *Strategic Management Journal, 20*, 195-204.

Iacobucci, D. (2010). Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology, 20*, 90-98.

International Society for the Scholarship of Teaching and Learning. (2004). Retrieved from https://www.issotl.com

Jöreskog, K. G., & Wold, H. O. A. (1982). The ML and PLS Techniques For Modeling with Latent Variables: Historical and Comparative Aspects. In H. O. A. Wold & K. G. Jöreskog (Eds.), *Systems Under Indirect Observation, Part I* (pp. 263-270). Amsterdam: North-Holland.

Joseph F. Hair, J., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis*. Upper Saddle River, NJ, USA: Pearson Prentice Hall.

Kaynak, H. (2003). The relationship between total quality management practices and their effects on firm performance. *Journal of Operations Management*, *21*, 405-435.

Kiley, M., & Wisker, G. (2009). Threshold concepts in research education and evidence of threshold crossing. *Higher Education Research & Development, 28*(4), 431–441.

Kim, D., & Kim, S.-K. (2012). Comparing patterns of component loadings: Principal Component Analysis (PCA) versus Independent Component Analysis (ICA) in analyzing multivariate non-normal data. *Behavior Research Methods, 44*(4), 1239-1243.

Land, R. (2015). Facilitating the Academy through Threshold Concepts and Troublesome Knowledge. In *On the Facilitation of the Academy* (Vol. 1, pp. 17-29). Rotterdam: Sense Publishers.

Land, R., Cousin, G., Meyer, J. H. F., & Davies, P. (2005). *Threshold concepts and troublesome knowledge (3): implications for course design and evaluation.* Paper presented at the Improving Student Learning - diversity and inclusivity, Oxford.

Land, R., Meyer, J. H. F., & Baillie, C. (2010). Editors' Preface: Threshold Concepts and Transformational Learning. In J. H. F. Meyer, R. Land, & C. Baillie (Eds.), *Threshold Concepts and Transformational Learning* (pp. ix-xlii). Rotterdam: Sense Publishers.

Land, R., Rattray, J., & Vivian, P. (2014). Learning in the liminal space: a semiotic approach to threshold concepts. *Higher education, 64*, 199-217.

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33*(1), 159-174.

Le Pallec, X., Marvie, R., Rouillard, J., & Tarby, J.-C. (2010). *A Support to Multi-devices Web Application.* Paper presented at the Proceedings of the 23th annual ACM Symposium on User Interface Software and Technology (UIST '10), New York.

Li, X. (2006). *Supportive Leadership, Learning Capability, IT Support Capability, Power and Value Appropriation in IOS Supply Chain Network Context (Unpublished doctoral dissertation)*. Retrieved from Toledo, OH, USA:

Lowry, P. B., & Gaskin, J. (2014). Partial Least Squares (PLS) Structural Equation Modeling (SEM) for Building and Testing Behavioral Causal Theory: When to Choose It and How to Use It. *IEEE Transactions on Professional Communication*, *57*(2), 123-146.

Luker, P. A. (1994). *There's more to OOP than syntax!* Paper presented at the SIGCSE '94: Proceedings of the twenty-fifth SIGCSE symposium on Computer science education.

Lunt, B. M., Ekstrom, J. J., Gorka, S., Hislop, G., Kamali, R., Lawson, E., . . . Reichgelt, H. (2008). *Information Technology 2008: Curriculum Guidelines for Undergraduate Degree Programs in Information Technology*. Retrieved from

Maciocha, A. (2014). *Teaching Research Methods: Threshold Concepts*, Carlow, Ireland.

Male, S. A., & Baillie, C. A. (2011). *Engineering Threshold Concepts.* Paper presented at the WEE2011, Lisbon, Portugal.

Malhotra, M. K., & Grover, V. (1998). An assessment of survey research in POM: from constructs to theory. *Journal of Operations Management*, 407-425.

Marton, F., Runesson, U., & Tsui, A. B. M. (2004). The Space of Learning. In *Classroom discourse and the space of learning* (pp. 3-43). New York, NY: Routledge.

McCartney, R., Boustedt, J., Eckerdal, A., Moström, J. E., Sanders, K., Thomas, L., & Zander, C. (2009). Liminal spaces and learning computing. *European Journal of Engineering Education*, *34*(4), 383-391.

McCartney, R., Eckerdal, A., Moström, J. E., Sanders, K., & Zander, C. (2007). *Successful Students' Strategies for Getting Unstuck.* Paper presented at the ITiCSE'07, Dundee, Scotland.

Meyer, J. H. F. (2016). Threshold concepts and pedagogic representation. *Education + Training, 58*(5), 463-475.

Meyer, J. H. F., & Land, R. (2003). Threshold Concepts and Troublesome Knowledge: linkages to ways of thinking and practising within the disciplines. In *Improving Student Learning – Ten Years On* (pp. 1-15). Oxford, UK: OCSLD.

Meyer, J. H. F., & Land, R. (2005). Threshold concepts and troublesome knowledge (2): Epistemological considerations and a conceptual framework for teaching and learning. *Higher education*, *49*, 373-388.

Meyer, J. H. F., & Land, R. (2006a). Editors' preface. In R. Land & J. H. F. Meyer (Eds.), *Overcoming Barriers to Student Understanding: Threshold concepts and troublesome knowledge* (pp. xiv-xxiii). London and New York: Routledge.

Meyer, J. H. F., & Land, R. (2006b). Threshold concepts and troublesome knowledge - An introduction. In J. F. H. Meyer & R. Land (Eds.), *Meyer and Land Overcoming Barriers to Student Understanding - Threshold concepts and troublesome knowledge* (pp. 3-18). London and New York: Rouyledge.

Meyer, J. H. F., Land, R., & Davies, P. (2008). Threshold concepts and troublesome knowledge (4): Issues of variation and variability. In R. Land, J. H. F. Meyer, & J. Smith (Eds.), *Threshold Concepts within the Disciplines* (pp. 59-74). Rotterdam: Sense Publishers.

Miller, C. S., & Connolly, R. (2015). Introduction to the special issue on web development. *ACM Transactions on Computing Education*, *15*(1), 1:1-1:5.

Miller, C. S., & Dettori, L. (2008). *Employers' Perspectives on IT Learning Outcomes.* Paper presented at the SIGITE '08 Proceedings of the 9th ACM SIGITE Conference on Information Technology Education, Cinxinnati, OH, USA.

Moore, G. C., & Benbasat, I. (1991). Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. *Information Systems Research, 2*(3), 192-222.

Morgan, P. K. (2015). *Foundational Assumptions in Threshold Concepts and Information Literacy.* Paper presented at the Creating Sustainable Community: The Proceedings of the ACRL 2015 Conference, Portland.

Morrell, A., & O'Connor, M. A. (2002). Introduction. In *Expanding the boundaries of transformative learning: Essays on theory and praxis* (pp. 280). New York: Palgrave.

Morris, R. D. (2011). Web 3.0: Implications for Online Learning. *TechTrends, 55*(1), 42-46.

Moström, J. E., Boustedt, J., Eckerdal, A., McCartney, R., Sanders, K., Thomas, L., & Zander, C. (2008). *Concrete examples of abstraction as manifested in students' transformative experiences.* Paper presented at the Proceedings of the Fourth international Workshop on Computing Education Research (ICER '08), Sydney, Australia.

Moström, J. E., Boustedt, J., Eckerdal, A., McCartney, R., Sanders, K., Thomas, L., & Zander, C. (2009). *Computer science student transformations: changes and causes.* Paper presented at the Proceedings of the 14th annual ACM SIGCSE conference on Innovation and Technology in Computer Science Education, Paris.

Murugesan, S. (2008). Web Application Development: Challenges and The Role of Web Engineering. In *Web Engineering: Modelling and Implementing Web Applications* (pp. 7-32). London: Springer.

Murugesan, S., & Deshpande, Y. (2002). *Meeting the Challenges of Web Application Development: The Web Engineering Approach (tutorial).* Paper presented at the International Conference on Software Engineering (ICES), Orlando, FL.

Nahm, A. Y., Rao, S. S., Solis-Galvan, L. E., & Ragu-Nathan, T. S. (2002). The Q-Sort Method: Assessing Reliability And Construct Validity Of Questionnaire Items At A Pre-Testing Stage. *Journal of Modern Applied Statistical Methods*, *1*(1), 114-125.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling Procedures: Issues and Applications*. Thousand Oaks, CA, USA: Sage Publications.

Nevo, B. (1985). Face Validity Revisited. *Journal of Educational Measurement, 22*(4), 287-293.

Norman, D. A., & Draper, S. W. (1986). *User-Centered System Design: New Perspectives on Human-Computer Interaction*. Hillsdale, NJ: Lawrence Earlbaum Associates.

O'Reilly, T. (2007). What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communicatioins & Strategies, 65*, 17-37.

O’Rourke, N., & Hatcher, L. (2013). *A Step-by-Step Approach to Using SAS® for Factor Analysis and Structural Equation Modeling*. Cary, NC: SAS Institute.

Or-Bach, R., & Lavy, I. (2004). Cognitive activities of abstraction in object orientation: an empirical study. *SIGCSE Bulletin, 36*(2), 82-86.

Park, E.-J., & Light, G. (2010). Identifying a potential threshold concept in nanoscience and technology. In J. H. F. Meyer, R. Land, & C. Baillie (Eds.), *Threshold Concepts and Transformational Learning* (pp. 260-279). Rotterdam: Sense Publishers.

Park, T. H., & Wiedenbeck, S. (2011). *Learning web development: Challenges at an earlier stage of computing education.* Paper presented at the Proceedings of the Seventh International Workshop on Computing Education Research (ICER'11), New York, NY.

Parry, S. (2019). Fit Statistics commonly reported for CFA and SEM. Retrieved from https://www.cscu.cornell.edu/news/Handouts/SEM_fit.pdf

Patrick, H., & Fields, Z. (2017). A Need for Cyber Security Creativity. In *Collective Creativity for Responsible and Sustainable Business Practice* (pp. 42-61): IGI Global.

Perkins, D. (1999, November 1). The many faces of constructivism. *Educational Leadership, 3,* 6-11.

Perkins, D. (2006). Constructivism and troublesome knowledge. In J. F. H. Meyers & R. Land (Eds.), *Overcoming Barriers to Student Understanding - Threshold concepts and troublesome knowledge* (pp. 33-47). London and New York: Routledge.

Perkins, D. (2008). Beyond Understanding. In R. Land, J. H. F. Meyer, & J. Smith (Eds.), *Threshold Concepts within the Disciplines* (pp. 3-19). Rotterdam: Sense Publishers.

Perreault, W. D., & Leigh, L. E. (1989). Reliability of Nominal Data Based on Qualitative Judgments. *Journal of Marketing Research, 26*, 135-148.

Qrunfleh, S., & Tarafdar, M. (2014). Supply Chain Information Systems Strategy: Impacts on Supply Chain Performance and Firm Performance. *International Journal of Production Economics*, *147*, 340-350.

Rhem, J. (2013). Before and after students "get it": threshold concepts. *Teaching Commons - Teaching Talk.* Retrieved from https://teachingcommons.stanford.edu/teaching-talk/and-after-students-get-it-threshold-concepts

Ringle, C. M., Sarstedt, M., & Straub, D. W. (2012). Editor's Comments: A Critical Look at the Use of PLS-SEM in MIS Quarterly. *MIS Quarterly, 36*(1), iii-xiv.

Ringle, C. M., Wende, S., & Becker, J.-M. (2015). SmartPLS 3. (Version 3.3.3): Boenningstedt: SmartPLS GmbH. Retrieved from http://www.smartpls.com

Ross, P. M., Taylor, C. E., Hughes, C., Whitaker, N., Lutze-Mann, L., & Tzioumis, V. (2010). *Threshold Concepts: Challenging the Way We Think, Teach, and Learn in Biology and Science.* Paper presented at the Proceedings of the 16th UniServe Science Annual Conference, Sydney, Australia.

Rountree, J., & Rountree, N. (2009). *Issues Regarding Threshold Concepts in Computer Science.* Paper presented at the Proceedings of the 11th Australasian Computing Education Conference (ACE 2009), Wellington, New Zealand.

Rowbottom, D. P. (2007). Demystifying Threshold Concepts. *Journal of Philosophy of Education, 41*(2), 263-270.

Russo, N. L. (2000). Developing Applications for the Web: Exploring Differences Between Traditional and World Wide Web Application. In M. Khosrowpour (Ed.), *Managing Web-enabled Technologies in Organizations : A Global Perspective.* (pp. 23-35). Hershey, PA, USA: IGI Publishing.

Sabin, M., Alrumaih, H., Impagliazzo, J., Lunt, B., Zhang, M., Byers, B., . . . Viola, B. (2017). *Information Technology Curricula 2017*. Retrieved from

Sanders, K., Boustedt, J., Eckerdal, A., McCartney, R., Moström, J. E., Thomas, L., & Zander, C. (2012). *Threshold Concepts and Threshold Skills in Computing.* Paper presented at the Proceedings of the ninth annual international conference on International computing education research, Auckland, New Zealand.

Sanders, K., & McCartney, R. (2016). *Threshold concepts in computing: past, present, and future.* Paper presented at the Proceedings of the 16th Koli Calling International Conference on Computing Education Research (Koli Calling '16), Koli, Finland.

Sanders, N. R. (2008). Pattern of information technology use: The impact on buyer-supplier coordination and performance. *Journal of Operations Management*, *26*, 349-367.

Schermelleh-Engel, K., Moosbrugger, H., & ;Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online, 8*(2), 23-74. Retrieved from http://www.mpr-online.de

Sendall, P., Ceccucci, W., & Peslak, A. (2008). Web 2.0 matters: An analysis of implementiong Web 2.0 in the classroom. *Information Systems Education Journal, 6*(64), 3-17.

Shinners-Kennedy, D. (2008). The Everydayness of Threshold Concepts: State as an Example from Computer Science. In R. Land, J. H. F. Meyer, & J. Smith (Eds.), *Threshold Concepts within the Disciplines* (pp. 119-128). Rotterdam: Sense Publishers.

Shinners-Kennedy, D., & Fincher, S. A. (2013). *Identifying threshold concepts: from dead end to a new direction.* Paper presented at the ICER '13 Proceedings of the ninth annual international ACM conference on International computing education research, San Diego, California.

Simonetto, A. (2012). Formative and Reflective Models: State of the Art. *Electronic Journal of Applied Statistical Analysis, 5*(3), 452-457.

Simsek, Z., & Veiga, J. F. (2001). A Primer on Internet Organizational Surveys. *Organizational Research Methods, 4*(3), 218-235.

Sorva, J. (2010). *Reflections on Threshold Concepts in Computer Programming and Beyond*. Paper presented at the Koli Calling '10 Proceedings of the 10th Koli Calling International Conference on Computing Education Research, Koli, Findland.

Statistics, U. S. B. o. L. (Producer). (2018, May). Occupational Employment and Wages, May 2018 15-1134 Web Developers. *Bureau of Labor Statictics Occupational Employment Statistics*. Retrieved from https://www.bls.gov/oes/current/oes151134.htm

Steiner, H. (2018, 7/31/2018). Teaching Journals Directory. Retrieved from https://cetl.kennesaw.edu/teaching-journals-directory

Stephanidis, C. (2001). User Interfaces for All: New Perspectives into Human-Computer Interaction. In C. Stephanidis (Ed.), *User Interfaces for All - Concepts, Methods, and Tools* (pp. 3-17). Mahwah, NJ: Lawrence Erlbaum Associates.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B, 36*(1), 111-147.

Sullivan, G. M., & Artino Jr, A. R. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education, 5*(3), 541-542.

Teaching and Learning Research Programme. (2009, 9). *Phase II Research Project: Enhancing Teaching-Learning Environments in Undergraduate Courses*. Retrieved from http://www.tlrp.org/proj/phase11/phase2c.html

Townson, L., Lu, S., Hofer, A. R., & Brunetti, K. (2015). What's the Matter with Threshold Concepts? *ACRLog*. Retrieved from http://acrlog.org/2015/01/30/whats-the-matter-with-threshold-concepts/

Turner, V. (1969). *The Ritual Process: Structure and Anti-Structure*. London: Routledge and Keegan Paul.

van Gennep, A. (1960). *The Rites of Passage: a classic study of cultural celebrations*. London: outledge and Kegan Paul.

Vockell, E. L. (1983). *Educational Research*. New York: Macmillan.

W3C (Producer). (2013, 12 11). W3C Semantic Web Activity Home Page. *W3C Semantic Web Activity*. Retrieved from http://www.w3.org/2001/sw/

Walker, G. (2013). A cognitive approach to threshold concepts. *Higher education, 65*(2), 247-263.

Wang, Y. D., & Zahadat, N. (2009). *Teaching Web Development in the Web 2.0 Era.* Paper presented at the SIGITE'09: Proceedings of the 10th ACM conference on SIG-information technology education, New York.

Wegner, E. (1998). Communities of Practice: Learning as a Social System. *Systems Thinker, 9*(5), 2-3.

Weresh, M. H. (2014). Stargate: Malleability as a Threshold Concept in Legal Education. *Journal of Legal Education, 63*(4), 689-728.

Weston, R., & Gore Jr., P. A. (2006). A Brief Guide to Structural Equation Modeling. *The Counseling Psychologist, 34*(5), 719-751.

Wilkinson, L. (2014). The problem with threshold concepts. *Sense and Reference Blog*. Retrieved from https://senseandreference.wordpress.com/2014/06/19/the-problem-with-threshold-concepts/

Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine, 8*(3).

Wilson, A., Åkerlind, G., Francis, P., Kirkup, L., McKenzie, J., Pearce, D., & Sharma, M. D. (2010). *Measurement uncertainty as a threshold concept in physics.* Paper presented at the Proceedings of the 16th UniServe Science Annual Conference, Sydney, Australia.

Wong, K. K.-K. (2013). Partial Least Squares Structural Equation Modeling (PLS-SEM) Techniques Using SmartPLS. *Marketing Bulletin, 24*(1), 1-32.

Wright, K. B. (2005). Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services. *Journal of Computer-Mediated Communication*, *10*(3).

Yip, J., & Raelin, J. A. (2012). Threshold concepts and modalities for teaching leadership practice. *Management Learning*, *43*(3), 333-354.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology, 9*(2), 79-94.

Zwaneveld, B., Perrenet, J., & Bloo, R. (2016). Discussion of Methods for Threshold Research and an Application in Computer Science. In M. A. Peters (Ed.), *Educational Futures - Rethinking Theory and Practice* (Vol. 68, pp. 269-284): Sense Publishers.